مقارنة طرق كيرنيل والمئينات وطرق نظرية استجابة الفقرة عند استخدام تصميم الفقرات المشتركة في دقة معادلة درجات الاختبارات متعددة الحدود

إعداد:

يوسف عبد العاطي محمد المحروق

المشرف

الأستاذ الدكتور خليل محمد عليان

قدمت هذه الأطروحة استكمالا لمتطلبات الحصول على درجة الدكتوراه في

علم النفس التربوي /القياس والتقويم

كلية الدراسات العليا

الجامعة الأردنية

ايار ٢٠١١

الجامعة الأردنية

نمـــوذج تفـــويض

أنا يوسف عبد العاطي محمد الجرو أفوض الجامعة الأردنية بتزويد نسخ مـن أطروحتي للمكتبات أو المؤسسات أو الهيئات أو الأشخاص عند طلبها.

التوقيع:

التاريخ: ٢٢/٦/٢٠١١

قرار لجنة المناقشة

نوقشت هذه الأطروحة " مقارنة طرق كيرنيل والمئينات وطرق نظرية استجابة الفقرة عند استخدام تصميم الفقرات المشتركة في دقة معادلة درجات الاختبارات متعددة الحدود." وأجيزت بتاريخ 28 / 2011/4

أعضاء لجنة المناقشـــــة	التوقيـــــــع
الأستاذ الدكتور خليل عليان/ مشرفاً أستاذ القياس والتقويم/ الجامعة الأردنية	....
الأستاذ الدكتور محمد وليد البطش/ عضوا أستاذ القياس والتقويم/ الجامعة الأردنية	....
الأستاذ الدكتور عبد الرحمن عدس أستاذ القياس والتقويم / متقاعد	....
الأستاذ الدكتور عدنان العابد أستاذ مناهج وأساليب تدريس الرياضيات	....

الإهداء

إلى من رباني صغيرا ودعماني كبيرا

أمي وأبي

إلى من وصلني في زمن القطع ،وساندني في زمن الوحدة

والدي خضر يوسف شريم

إلى أول شخص دعم الحلم ، ورعاه بدقة ،إلى رأس الحكمة وعنوان الفكرة:

والدتي فاطمة سعادة.

إلى حبيبتي ورفيقة دربي، إلى من أعطت لحياتي معنى، إلى من تناولت حلمي بالرعاية والدعم ، وهي ذات ذكاء ثاقب وقلب ثابت وحكمة فطرية

حبيبتي وزوجتي مي شريم.

إلى وردات قلبي ومن ملأه حبا وحياة ، فلذات كبدي:

محمد ، احمد ، محمود.

إلى من له الجميل بدو الجميل أخي

انس شريم.

إلى من رافقوني في الطفولة ودعموا الحلم

إخوتي وأخواتي.

شكر وتقدير

الحمد لله حمدا ينبغي لجلال وجهه ، وعظيم قدره و سلطانه ، الذي يسر لي إتمام هذه الرسالة ، والصلاة والسلام على اشرف خلق الله ، سيدنا محمد وعلى آله وصحبه وسلم وبعد،،،

فان أي امة مهما بلغت من قوة وحصانة ومنعة ، لا يمكن أن تحافظ على ديمومة كينونتها ما لم تعترف بعظمة علمائها ، وقدرتهم على التغيير ، وتوليهم كل عناية واحترام ؛لما لهم من اثر فعال في نقل الأمة من سني التخلف إلى سني الرقي والتقدم والازدهار، وعطفا عليه فإنني أرى من واجبي أن أرد الفضل إلى أهله ، وان أقدم جزيل الشكر والعرفان ، لكل من وقف معي وقدم لي النصح والإرشاد في جميع مراحل الدراسة .

فخالص تقديري وعظيم مودتي إلى أستاذي الذي به افخر وأزهو أستاذي الأستاذ الدكتور خليل محمد عليان ، المشرف على هذه الأطروحة، لما قدم لي من خالص النصيحة، وسداد الرأي ، وإفادتي بملاحظاته الدقيقة ، وإرشاداته القيمة التي كان لها اكبر الأثر في إتمام هذه الأطروحة ، وفي إخراجها إلى حيز الوجود.

وأتقدم بالشكر والامتنان وخالص الاحترام: إلى أعضاء لجنة المناقشة الممثلة برئيس اللجنة الأستاذ الدكتور: خليل عليان، والأستاذ الدكتور محمد البطش عضواً، والأستاذ الدكتور عبد الرحمن عدس عضواً، والأستاذ الدكتور عدنان العابد عضواً، على تفضلهم بقبول مناقشة هذه الأطروحة ولما أبدوه من توجيهات ساعدت في إثراء هذه الأطروحة بالمعلومات القيمة والمفيدة.

وأتقدم بالشكر الجزيل لكل من الدكتور نزار اللبدي وطارق الدراغمة ومحمد حبش والمهندس احمد عرفات لما بذلوه من جهد برمجي شاق ، كان له اكبر الأثر في إتمام هذا العمل.

فهرس المحتويات

قائمة الجداول

قائمة الأشكال

مقارنة طرق كيرنيل والمئينات وطرق نظرية استجابة الفقرة عند استخدام تصميم الفقرات المشتركة في دقة معادلة درجات الاختبارات متعددة الحدود

إعداد

يوسف عبد العاطي محمد المحروق

إشراف

الأستاذ الدكتور خليل عليان

الملخص

هدفت هذه الدراسة إلى مقارنة طرق كيرنيل والمئينات وطرق نظرية استجابة الفقرة(IRT) عند استخدام تصميم الفقرات المشتركة في دقة معادلة درجات الاختبارات متعددة الحدود ،وذلك من خلال دراسة المتغيرات الآتية :

١- حجم العينة: استخدمت ثلاثة مستويات لحجم العينة وهي: ٢٠٠ مفحوص ، ٦٠٠مفحوص، ١٠٠٠ مفحوص ،وهذه الأحجام للعينات تعتبر مناسبة لطرق المعادلة باستخدام تصميم المجموعات العشوائية.

٢- طول الاختبار : استخدمت ثلاثة مستويات لطول الاختبار وهي ٣٠-٤٠-٦٠ فقرة.

٣- مستويات الصعوبة: استخدم مستويان من الصعوبة: التشابه في الصعوبة، واختلاف الصعوبة (اختلاف الصعوبة=٠,٠١ انحراف معياري).

تم التوصل إلى نتائج دقة المعادلة في ظل استخدام المتغيرات السابقة منفردة ومجتمعة ، حيث تم تقدير دقة المعادلة باستخدام معياري : الخطأ المعياري للمعادلة ، والجذر ألتربيعي لمتوسط الخطأ.

وبالتحديد فان هذه الدراسة حاولت الإجابة عن السؤالين الآتيين:

- هل تختلف طرق المعادلة: كيرنيل، المئينات،IRT ، من حيث الخطأ المعياري للمعادلة تحت كل ظرف من: طول الاختبار، حجم العينة، وتشابه مستويات الصعوبة؟

- هل تختلف طرق المعادلة : كيرنيل، المئينات،IRT، من حيث جذر متوسط مربعات الخطأ للمعادلة تحت كل ظرف من: طول الاختبار، حجم العينة، وتشابه مستويات الصعوبة؟

واختارت هذه الدراسة الطريقة المئينية وطريقة نظرية استجابة الفقرة وطريقة كيرنيل؛ لأنها أكثر الطرق شيوعاً في المعادلة؛ وذلك للكشف عن دقة معادلة الدرجات في الاختبارات مستخدمين تصميم المجموعات العشوائية ذات الفقرات المشتركة.

ولمقارنة طرق المعادلة الثلاث، تم توليد بيانات تجريبية باستخدام برمجية (Wingen2)، وذلك باستخدام ثلاثة متغيرات: طول الاختبار، حجم العينة، والتشابه في مستويات الصعوبة، وتم استخدام نظرية استجابة الفقرة لتوليد استجابات المفحوصين على فقرات الاختبار، حيث تم معادلة درجات الاختبارات باستخدام الدرجات الملاحظة في نظرية استجابة الفقرة كمعيار رئيسي للمعادلة، كذلك تم استخدام الخطأ المعياري للمعادلة والبواقي المعيارية لتقييم نتائج المعادلة.

أشارت النتائج أن طريقة المعادلة باستخدام نظرية استجابة الفقرة كانت أكثر الطرق دقة، ثم تلتها طريقة كيرنيل، وعند مختلف الظروف التي تم تناولها في هذه الدراسة ، كذلك أشارت النتائج أن حجم العينات الكبير يقلل من الخطأ المعياري للمعادلة، ويقلل كذلك من جذر متوسط مربع الخطأ.

الفصل الأول

مقدمة الدراسة وأهميتها

الفصل الأول

مقدمة الدراسة وأهميتها

المقدمة:

ظل العاملون في القياس النفسي يستخدمون مبادىء واسس النظرية الكلاسيكية في بناء الاختبارات والمقاييس باشكالها المختلفة، وتفسير الدرجات المتحققة عليها لفترة طويلة من الوقت ، على الرغم من جوانب القصور التي تعاني منها .

وقد حدد هامبلتون ولندن هذه الجوانب في النقاط الآتية:

١- اختلاف خصائص فقرات الاختبار باختلاف عينات المفحوصين المستخدمين في معايرة الفقرات؛ إذ أن خصائص فقرات الاختبار تعتمد على مستوى وتوزيع قدرات أفراد هذه العينات، فصعوبة الفقرات تكون أعلى مما هي في الواقع في حالة اختيار عينة تكون في مستوى قدرتها أعلى من مستوى قدرة الأفراد في المجتمع، كما تكون ادني مما يجب في حالة اختيار عينة تكون في مستوى قدرتها اقل من مستوى قدرة أفراد المجتمع.

٢- تكون معاملات التمييز أعلى في حالة كون توزيع أفراد العينة أكثر تباينا من توزيع أفراد المجتمع، والعكس صحيح.

٣- يختلف تقدير العلامة الكلية للمفحوصين باختلاف عينة الفقرات المقدمة لهم، إذ تعمل الفقرات الصعبة على خفض مستوى العلامة الكلية للمفحوصين.

٤- إن الاختبارات في النظرية الكلاسيكية لا تزودنا بتقدير دقيق للمفحوصين ذوي القدرة العالية وللمفحوصين ذوي القدرة المتدنية؛ وذلك لأن الاختبارات المبنية على أساس النظرية الكلاسيكية في القياس تأخذ في اعتبارها عند اختيار الفقرات أن تتلاءم مع الأفراد متوسطي القدرة (Hambelton &linden, 1982)

ولتلافي العيوب التي ظهرت في النظرية الكلاسيكية، ومن اجل الوصول إلى قياس أكثر موضوعية، جاءت النظرية الحديثة ، والتي اصطلح على تسميتها بنظرية الاستجابة للفقرة (Item Response Theory)، أو ما يسمى بالنظرية الحديثة في القياس .

لقد بين هامبلتون وسوامينثان Hambelton & Swaminathan (1985) مزايا رئيسيه لنظرية استجابة الفقرة منها:

١- بافتراض وجود عدد كبير من فقرات الاختبار التي تقيس السمة، فان تقدير قدرة الفرد يكون مستقلا عن عينة الفقرات التي تطبق عليه، أي أن تقديرات القدرة للأفراد متحررة من خصائص الفقرات المستخدمة في تقدير القدرة (Item free).

٢- بافتراض وجود عدد كبير من الأفراد، يكون تقدير معالم الفقرات مستقلا عن عينة الأفراد التي استخدمت في تقدير هذه المعالم (Sample free).

إن نظرية الاستجابة للفقرة تقوم على افتراضات أساسية ذكرها هامبلتون و سوامينثيان (1991) Hambelton&Swaminathan وهي: افتراض أحادية البعد(Unidimensionality)، وافتراض الاستقلال الموضعي (Local Independence) والمطابقة لمنحني خصائص الفقرة(ICC) ، والتحرر من سرعة الأداء (Speediness).

كذلك تعتمد نظرية استجابة الفقرة على وجود عدد من النماذج الرياضية المستخدمة لمطابقة النموذج للبيانات التي من أهمها: النماذج اللوجستيه الأحادية والثنائية الثلاثية والرباعية.

إن النجاح في استخدام نماذج نظرية الاستجابة للفقرة يقوم عن مجموعة من المتطلبات منها: أحادية البعد للاختبار، مطابقة فقرات الاختبار للنموذج المستخدم ، حجم العينة المستخدمة، توفر البرمجيات الحاسوبية المناسبة، والطريقة المستخدمة في تقدير معالم الفقرات ومعلمة القدرة (Hambelton & Swaminathan, 1991).

لقد ظهرت عدد من التطبيقات العملية لنظرية استجابة الفقرة ومن هذه التطبيقات: الكشف عن التحيز في فقرات الاختبارات،وبناء بنوك الأسئلة، والقياس التكيفي، بناء الاختبارات وتحليلها وتقنينها، وكذلك معادلة الدرجات على صور الاختبار المختلفة.

يحتاج المعلمون عادة إلى تطبيق عدد من الاختبارات المتعادلة في محتواها وفي مستوى درجة صعوبتها، ويقصد بمعادلة صورتي الاختبار(Test Equating) تحويل نظام وحدة القياس (الدرجات) الخاص بإحدى الصورتين إلى نظام وحدات القياس (الدرجات) الخاص بالصورة الأخرى، بحيث تصبح القياسات المستخدمة في درجات كل من الصورتين

متعادلة بعد إجراء هذا التحويل، حيث تكون العلامة المتحققة على أي منها مناظرة لموقع الفقرة على متصل قياس القدرة بدلالة نفس وحده القياس مهما كان ذلك الموقع.

تعتمد معادلة الاختبارات على وجود نوعين من أنواع المعادلة وهما: المعادلة الأفقية (Horizontal equating) والمعادلة الرأسية (Vertical equating): ففي المعادلة الأفقية تتصف صور الاختبارات المراد معادلتها في المعادلة الافقية بان لها مستويات صعوبة متقاربة وتوزيعات القدرة لدى المفحوصين متشابهة ، ويكون الهدف من المعادلة هو تعديل الفروق الناتجة عن الاختلاف في مستويات الصعوبة بين صورتي الاختبار ، وتكون صور الاختبارات في المعادلة الافقية عبارة عن صور متعددة ومتوازنة من نفس الاختبار ، الا ان الفقرات في الصور المختلفة للاختبارات ليست متطابقة ، ولذا تستخدم المعادلة الافقية عندما يكون الهدف من تعدد صور الاختبار هو الحفاظ على سرية الاختبارات ، فمثلا عندما يكون الهدف هو قياس مستوى التحصيل لعدد كبير من الصفوف لطلاب الصف السادس في مجتوى دراسي محدد ،فان ذلك يستدعي وجود عدة صور من الاختبار ، لذا فان الاختبارين متشابهان في الصعوبة والمحتوى ويقدمان الى افراد عينة متشابهين في توزيع القدرة ، حيث يعد من الضروري وجود صور متعددة من الاختبار لضمان سريته ، كما يحدث في حالة اختبارات القبول في الكثير من الجامعات .
وبالتالي فانه من المفترض ان تكون صور الاختبار مختلفة في الفقرات ، ايضا من تطبيقات هذا النوع من المعادلة هو تطبيقات الانظمة التعليمية التي تسمح باعادة الاختبار في حالة الفشل ، وتكون الاعادة باستخدام صور مختلفة للاختبار تم معادلتها لتعطي درجات قابلة للمقارنة مع درجات صور الاختبار الاخرى (الحربي ، ٢٠٠٩)

اما المعادلة العمودية ، فتكون الاختبارات المراد معادلتها مختلفة الصعوبة ، وتوزيعات القدرة لدى المفحوصين مختلفة ، والقدرة او السمة تقع على متصل واسع المدى او متعدد المستويات، ويتضح ذلك في السمات التطورية ، ومن هذا المبدا تظهر الحاجة الى وجود عدة صور من الاختبار تقيس جميعها نفس السمة ولكن على مستويات مختلفة مقارنة بقدرات المفحوصين ، وبالتالي فان صور الاختبارات المراد معادلتها ذات مستويات مختلفة من حيث الصعوبة ، كذلك فان توزيعات القدرة لدى المفحوصين ستكون على مستويات متعددة ، وبالتالي اختلاف المفحوصين في توزيعات السمة المراد قياسها ، ويكون استخدام

المعادلة العمودية عند بناء مقياس واحد يسمح بالمقارنة بين عدة مستويات لدى المفحوصين في سمة ما يراد قياسها على عدة مستويات ، حيث ستختلف مستويات صعوبة الفقرات من مستوى الى آخر مع ان جميع صور هذه الاختبارات تقيس نفس السمة ، ويفيد ايضا هذا النوع من المعادلة في بطاريات الاختبارات التي تختلف في مستويات صعوبتها، وكذلك في مقارنة مستويات الطلاب في مراحل مختلفة ، كما يتم استخدام المعادلة العمودية في الدراسات الطولية ، حيث ان هناك اختلاف في محتوى اختبارات المراحل لكنها تتصف بالتركيز على نفس المفهوم العام ، ويتوقع ان يبدي الطلبة تحسنا في الاداء سنة تلو الاخرى، واستخدام عملية المعادلة العمودية يسمح بمقارنة الدرجات مباشرة على طول متصل او بعد مشترك، حيث تهدف هذه المعادلة الى انشاء مقياس مشترك لجميع الصور .
ان الفارق ما بين نوعي المعادلة يعتمد على مدى الاختلاف في صعوبة الاختبارات المراد معادلتها ، فضلا عن انتماء المفحوصين الى صف دراسي واحد (أي يقيسان نفس مستوى القدرة) ، او انتمائهم الى صفوف مختلفة(أي يقيسان نفس مستوى القدرة)، او انتمائهم الى صفوف مختلفة (أي يقيسان مستويات مختلفة من القدرة).

يعتبر لورد(1980) Lord صور الاختبار متكافئة عندما لا يشكل تطبيق أي اختبار منها أي فارق في أداء المفحوصين، حيث أن الاختبارين (X ، y) لا يمكن أن يعتبرا متكافئين إلا إذا كانت القدرة المستنتجة للفرد هي ذاتها، سواء تقدم الفرد للاختيار X أم للاختبار y (Lord ,1980, P117).

لا تتم المعادلة بشكل عشوائي، فهناك مجموعه من الشروط لابد من توافرها وهي:

١- أن تقيس الاختبارات نفس القدرة (Same Ability).

٢- أن يتحقق شرط التساوي في توزيع الدرجات (Equity)؛ أي يكون التوزيع التكراري الشرطي(Conditional Frequency Distribution) للعلامات على الصورة الأولى بعد التحويل هو التوزيع التكراري الشرطي نفسه للعلامات على الصورة الثانية، وذلك لكل مجموعه من المفحوصين.

٣- أن لا تتغير المعادلة بين علامات صورتي الاختبار بتغير المجموعة التي أجابت عن صورتي الاختبار وهذا ما يسمى بالثبات عبر المجموعات.

٤- أن يتحقق التماثل (Symmetry)؛ أي عدم تغير المعادلة بين صورتي الاختبار بغض النظر عن أيهما اعتبر الأساس في المعادلة. أي اذا تمت معادلة النموذج (X) الى نموذج (y) فاننا سنحصل على نفس النتائج في حال تم معادلة النموذج (y) الى نموذج (X) (Peterson,et al, 1989)

عند الحديث عن معادلة الصور المختلفة للاختبارات توجد ثلاث قضايا مترابطة و متداخلة يجب أخذها بعين الاعتبار وهي:

١- الأساس النظري الذي يتم بموجبه التناظر والمقابلة بين العلامات في الصور المختلفة للاختبار، فهناك نظريتان رئيسيتان يتم من خلالهما اشتقاق طريقة المعادلة وهما: النظرية التقليدية والنظرية الحديثة في القياس.

٢- الحاجة إلى تصميم لجمع البيانات المعادلة، فهناك عده أنواع من التصاميم لجمع البيانات حيث تتدخل عدة اعتبارات عملية في اختيار التصميم المناسب.

٣ – الطرق الإحصائية التي ستستخدم في تقدير العلامة المناظرة، مثل هل سيتم استخدام الطريقة الخطية أم الطريقة المئينية؟ (kolen, 1981)

مشكلة الدراسة وأسئلتها:

هدفت هذه الدراسة إلى مقارنة طريقة كيرنيل، المئينات، IRT، و تحت ظروف: البيانات متعددة الحدود، الفقرات المشتركة، حجم العينة، طول الاختبار، وتجانس مستوى الصعوبة. واختيرت هذه الطرق كموضوع لهذه الدراسة كونها الأكثر شيوعا واستخداما.

وبعبارة أخرى فان الغرض الأساسي من الدراسة هو: التحقق من دقة معادلة درجات الاختبارات ذات الفقرات المشتركة للمجموعات المتكافئة باستخدام عدة طرق لمعادلة درجات الاختبارات. ويمكن صياغة مشكلة الدراسة بالتساؤل الآتي:

هل تختلف معادلة الدرجات على صور الاختبار باختلاف طريقة المعادلة المستخدمة؟

وبالتحديد فان هذه الدراسة حاولت الإجابة عن السؤالين الآتيين:

- هل تختلف طرق المعادلة: كيرنيل، المئينات،IRT ، من حيث الخطأ المعياري للمعادلة تحت كل ظرف من: طول الاختبار، حجم العينة، وتشابه مستويات الصعوبة؟

- هل تختلف طرق المعادلة : كيرنيل، المئينات،IRT، من حيث جذر متوسط مربعات الخطا للمعادلة تحت كل ظرف من: طول الاختبار، حجم العينة، وتشابه مستويات الصعوبة؟

واختارت هذه الدراسة الطريقة المئينية وطريقة نظرية استجابة الفقرة وطريقة كيرنيل باعتبارها أكثر الطرق شيوعا في المعادلة؛ وذلك للكشف عن دقة معادلة الدرجات في الاختبارات مستخدمين تصميم المجموعات العشوائية ذات الفقرات المشتركة.

أهمية الدراسة:

تكتسب الدراسة أهميتها من خلال استخدامها البيانات متعددة الحدود في تصميم فقرات الجذع المشترك ، وذلك لمقارنة طرق كيرنيل، المئينات،IRT،تحت ظروف: حجم العينة، طول الأختبار، وتجانس مستوى الصعوبة. ويمكن تلخيص أهمية الدراسة في الجوانب الآتية:

الأهمية النظرية

- الإسهام في إلقاء الضوء على أفضل طريقة لمعادلة درجات الاختبارات من بين الطرق المستخدمة في هذه الدراسة، بحيث تسهم في إعطاء صورة واضحة للمتخصصين في مجال بناء الاختبارات لاختيار طريقة المعادلة التي تتواءم مع طبيعة الاختبار الذي يتم بناؤه للحصول على نتائج تمتاز بالشمولية والدقة.
- تبرز أهمية الدراسة في أن معظم الدراسات السابقة كانت تتم باختيار طريقة أو اثنتين من طرق المعادلة، بينما في هذه الدراسة سيتم استخدام ثلاث طرق للمعادلة للوقوف على أكثرها دقة في معادلة درجات الاختبارات.

الأهمية العملية

- توفير معلومات وإرشادات تساعد المهتمين في بناء الاختبارات لانتقاء أفضل الطرق في معادلة الدرجات وتطبيقها بكل يسر وسهولة.

التعريفات الإجرائية:

معادلة درجات الاختبارات: هو إجراء إحصائي يتم فيه تحويل سلم الدرجات على أحد الاختبارات إلى سلم الدرجات على الاختبار الآخر، بحيث يمكن معرفة درجة الفرد على أحد الاختبارات إذا علمنا درجته على الاختبار الآخر.

طرق معادلة الاختبار: عبارة عن إجراءات رياضية تم تطويرها لتحقيق إجراءات المعادلة.

اختبار الجذع المشترك: هو اختبار مكون من مجموعة من الفقرات المشتركة بين اختبارات الصفوف المختلفة.

دقة المعادلة: هو أسلوب إحصائي يستخدم للتأكد من مدى فاعلية المعادلة باستخدام اختبار الجذع المشترك.

البيانات متعددة الحدود : وهي فقرات اختباريه تكون الاستجابة عليها متعددة (١-٢-٣-٤-٥) حيث يكلف المفحوص بتحديد الاستجابة التي يراها مناسبة.

الفصل الثاني

الإطار النظري والدراسات السابقة

الفصل الثاني

الإطار النظري والدراسات السابقة

أولا :الإطار النظري:

تختلف نماذج الاختبارات غالبا في صعوبتها، وفي المفحوصين الذين سيطبق عليهم هذه النماذج، ولان هؤلاء المفحوصين سيتقدمون لنماذج سهلة أو صعبة من الاختبار ، فان هذا يؤدي إلى وجود أفضلية أو عدم أفضلية لهم في هذه الاختبارات نتيجة لاختلاف هذه النماذج في صعوبتها، وحتى يتم تعديل هذه الاختلافات في صعوبة النماذج، تستخدم المعادلة بحيث يمكن أن تستخدم الدرجات على نماذج الاختبار المختلفة بشكل تبادلي(Kolen, Brenann, 2004). فإذا كانت المعادلة ناجحة، فان المفحوصين سيكون لهم نفس الدرجات المتوقعة، بغض النظر عن أي نموذج تقدموا له، فالمعادلة تعمل على تحقيق عدالة الاختبار، وإزالة آثار التدريب، ومقارنة أداء المفحوصين عبر نماذج الاختبار المختلفة.

ومن الامور التي يجب مراعاتها عند اجراء المعادلة:

١- تحديد تصميم جمع المعلومات المناسب.

٢- التعريف الاجرائي للمعادلة(هل سيتم استخدام النظرية الكلاسيكية ،ام نظرية استجابة الفقرة).

٣- تحديد طريقة اجراء المعادلة بين صور الاختبار المختلفة .

٤- دقة اجراء المعادلة.(Kolen, Brenann, 2004)

تعتبر المعادلة إحدى الإجراءات المستخدمة لتحقيق عدالة الدرجات على الاختبارات. فعندما يتم تطبيق اختبارات متعاقبة ومتسلسلة، فان المعادلة تصبح ضرورية لوضع الدرجات في الاختبارات المختلفة على نفس السلم. وذلك نظرا لان استخدام وتطبيق الاختبار بشكل متكرر، فان فقرات الاختبار يمكن أن تصبح فقرات سهلة ومعروفة للمفحوصين المستقبليين؛ لذلك يتم إعداد نماذج متعددة للاختبارات في العديد من برامج الاختبارات. وعلى الرغم من وجود النماذج المتعددة للاختبارات والتي تم بناؤها بالاعتماد على نفس الخصائص (نفس المحتوى، ونفس مستوى الصعوبة)، إلا أن نماذج الاختبار لا تكون متكافئة بالضبط؛ ولهذا

السبب فان بعض المفحوصين الذين يأخذون الاختبار الأسهل سيكون لهم أفضلية عن أولئك الذين يأخذون الاختبار الأصعب. لذلك إذا تمت المعادلة بشكل مثالي، فان المفحوصين سوف يحصلون على نفس الدرجات بغض النظر عن أي اختبار يتم التقدم له (Kolen , Brenann,2004).

تعريف المعادلة(Definition of Equating):

المعادلة: هي إجراء إحصائي يتم فيه تعديل الاختلاف في مستوى الصعوبة من نموذج إلى آخر. ونتيجة لهذا التعريف ظهر جدل واسع بين علماء القياس حول ما هو نوع الإجراء الإحصائي المميز أو الظروف التي ستكون مقنعة إذا تم إجراء المعادلة؟

اقترح لورد (1980) Lord أربع خصائص للمعادلة كشروط مسبقة لاجراء المعادلة: نفس القدرة، خاصية العدالة، خاصية التماثل، وخاصية الثبات. ومن خلال هذه الخصائص فان معظم علماء القياس أيدوا خاصيتي الثبات والتماثل، ولكنهم اختلفوا حول خاصيتي العدالة ونفس القدرة.

تم توظيف تعريفين للمعادلة: تعريف معادلة الدرجات الحقيقية وتعريف معادلة الدرجات الملاحظة. فقد عرف لورد ووينجيرسكي (1984) Lord,Wingersky المعادلة عن طريق وصف خاصية العدالة حيث قالا(لأي مجموعة من جميع المفحوصين عند مستوى القدرة المعطاة، توزيع الدرجات بعد المعادلة هو نفسه بغض النظر عن أي اختبار تم إعطاؤه). من خلال هذا التعريف يتبين أن خاصية العدالة تعرف بالاعتماد على التوزيع الشرطي للدرجات الملاحظة المعطاة من الدرجات الحقيقية.

إن جملة(أي مجموعة من المفحوصين) التي أشار إليها لورد (Lord) في تعريفه للمعادلة تبين أن المعادلة يجب أن تكون نفسها بغض النظر عن المجموعات في المجتمع المستهدف.

ان تعريف الدرجات المشاهدة يتبع من خاصية المعادلة المئينية، فقد عرف آنكوف(1971) Angoff المعادلة كما يأتي: (تعتبر الدرجتان متكافئتان إذا كانت الرتب المئينية لأي مجموعة معطاة هي متساوية).

يمكن تصنيف طرق المعادلة إلى ثلاث فئات:

١. طرق معادلة الدرجات المشاهدة الكلاسيكية.

٢. طرق معادلة الدرجات الملاحظة حسب نظرية استجابة الفقرة((IRT)).

٣. طرق معادلة الدرجات الحقيقية حسب نظرية استجابة الفقرة((IRT)).

إجراءات المعادلة:

إن القيام بإجراءات المعادلة يعتمد على كل من: تصميم جمع المعلومات لاغراض المعادلة، وطريقة المعادلة، ويتضمن تصميم المعادلة عددا من التصاميم وهي:

أولا: تصميم المجموعات العشوائية المتوازنة(Counter-Balanced Design)

المجموعة العشوائية المتوازنة أو ما يسمى مربع لاتين (Latin Square) غالبا يستخدم في التصاميم التجريبية لإلغاء اثر التدريب، وفي مجال الاختبارات التربوية يستخدم تصميم المجموعات العشوائية المتوازنة(Counter-Balanced Design) لاختبار فاعلية البرامج (Akour,2006).

في هذا النوع من التصميم تأخذ مجموعتين مستقلتين من المفحوصين اختبارين متطابقين: نموذج (X) ونموذج (Y) في ترتيب مختلف، حيث يتم تقسيم المفحوصين عشوائيا إلى مجموعتين متساويتين وتطبق الصورة الأولى للاختبار على إحدى المجموعتين، وتطبق الصورة الثانية على المجموعة الثانية، ثم تأخذ كل مجموعة الصورة التي لم تأخذها في المرة الأولى مباشرة بعد أخذها الاختبار السابق على أن يكون الزمن المحدد لتقديم صورتي الاختبار متساويا للمجموعتين، وان تكون نسخ الاختبار معدة بشكل متسلسل.

ثانيا: تصميم المجموعة المتكافئة(Equivalent Group Design)

في تصميم المجموعة المتكافئة(Equivalent Group Design) يتم اختيار مجموعتين مستقلتين عشوائيا من مجتمع معين من المفحوصين، ثم تأخذ إحدى المجموعتين عشوائيا إحدى صور الاختبارين المتطابقين وتأخذ المجموعة الأخرى الصورة الثانية.

في هذا التصميم يتم تجاوز تعرض نفس المفحوص للإجابة على صورتي الاختبار وبالتالي عدم تأثر المفحوص بعوامل مثل التعلم والتدريب والتعب. ومن الطرق التي يتم استخدامها غالبا لتعيين النماذج عشوائيا على المفحوصين الطريقة التصاعدية، فعلى سبيل المثال إذا كان لدينا نموذجين من الاختبار وأردنا معادلتهما (النموذج الجديد والنموذج القديم)، فان كتيبات الاختبار للنموذجين يتم وضعهما في حزم متعاقبة ومتسلسلة (النموذج الجديد، النموذج القديم، النموذج الجديد، النموذج القديم، .....وهكذا) وفي أثناء وقت تطبيق الاختبار يتم انجاز الاختبار كما يأتي:

المفحوص الأول يأخذ النموذج الجديد، المفحوص الثاني يأخذ النموذج القديم، المفحوص الثالث يأخذ النموذج الجديد.... وهكذا.

إن الطريقة التصاعدية تساعد على ضمان أن المفحوصين في الموقف الاختباري سيكونون متشابهين في مكان وزمان اخذ الاختبار، وفي كافة ظروف تطبيق الاختبار، بالإضافة إلى هذا فان الطريقة التصاعدية تضمن أن المفحوصين سيتم تقسيمهم بالتساوي على نموذجي الاختبار، لذلك إذا كان هناك اختلاف في معدل الأداء بين مجموعتي المفحوصين على النماذج، فان ذلك سيكون مؤشرا على اختلاف معدل الصعوبة بين النموذجين (Alina Von Daveir,et,al. 2004).

لهذا التصميم مجموعة من الايجابيات والتي نذكر منها: إلغاء عوامل التعب والتدريب، وتقليل وقت الاختبار لكون كل مفحوص يأخذ فقط نموذج واحد من الاختبار، ويكون قادرا على معادلة أكثر من اختبار جديد في نفس الوقت من خلال تضمين هذه النماذج الجديدة في العملية التصاعدية.

ثالثا: تصميم المجموعة المفردة(Single Group Design)

في تصميم المجموعة المفردة(Single Group Design)، يتم اختيار عينة عشوائية من المفحوصين من مجتمع المفحوصين، بحيث يأخذ جميع افراد هذه العينة الذين تم اختيارهم الاختبارين (x) و (y) في نفس الظروف.

بما أن الاختبارين متكافئان، وتم تطبيقهما على نفس المفحوصين، فان من المحتمل أن أداء المفحوصين على الاختبار الثاني سوف يتأثر بأدائهم على الاختبار الأول، وقد يكون هذا التأثر هو نتيجة لعوامل التدريب أو التعلم أو عوامل التعب.

من المألوف – مع ارتفاع الأداء على الاختبار– أن يكون النموذج (y) أسهل من النموذج (x)، ومن جانب آخر إذا اثر عامل التعب على أداء المفحوصين فانه سيترتب على ذلك أن يصبح النموذج (y) أكثر صعوبة من النموذج(x)؛ لان المفحوصين سيشعرون بالتعب عندما يعطون النموذج (y).(Alina Von Daveir,et,al. 2004).

رابعا: تصميم المجموعات غير المتكافئة ذات الفقرات المشتركة (Nonequivalent with Anchor Test)(NEAT):

في تصميم المجموعات غير المتكافئة ذات الفقرات المشتركة (Nonequivalent with Anchor Test)(NEAT)، تطبق الصورة الأولى للاختبار على المجموعة الأولى، و تطبق الصورة الثانية للاختبار على المجموعة الثانية، ثم يتم تطبيق الاختبار ذات الفقرات المشتركة (Anchor Test)على المجموعتين معا، ويمكن أن يكون الاختبار ذات الفقرات المشتركة داخليا (مجموعات جزئية من الفقرات يتم تضمينها في صورتي الاختبار المراد معادلتهما) أو خارجيا(اختبار ذو فقرات مشتركة تطبق في وقتين مختلفين على المجموعتين). (Alina Von Daveir,et,al. 2004).

خامسا: تصميم المجموعات المتكافئة ذات الفقرات المشتركة (Equivalent with Anchor Test)(EAT)

يشبه هذا التصميم تماما تصميم المجموعات غير المتكافئة ذات الفقرات المشتركة (Nonequivalent with Anchor Test)(NEAT)، والفرق الوحيد بينهما أن المجموعتين في تصميم المجموعات المتكافئة ذات الفقرات المشتركة يتم اختيارهما بالطريقة العشوائية لضمان تكافؤ المجموعتين.

نشأت فكرة الاختبار ذات الفقرات المشتركة كحل لعدم قدرة مصممي الاختبارات على بناء اختبارات متوازية (Parallel Tests) أو متكافئة تماما لغايات إجراء المعادلة، و الاختبار ذات الفقرات المشتركة عبارة عن مجموعة من الفقرات تمثل رابطا بين اختبارين أو أكثر بحيث يتم من خلالها إجراء عملية المعادلة، لذلك فقد تم إيجاد تصميمين جديدين للمعادلة باستخدام الاختبار ذات الفقرات المشتركة: تصميم المجموعات غير المتكافئة ذات الفقرات المشتركة وتصميم المجموعات المتكافئة ذات الفقرات المشتركة (الصمادي،٢٠٠٧).

وقد أشار بوديسكيو (Budescu,1985)المشار اليه في الصمادي (٢٠٠٧) إلى أن الاختبار ذات الفقرات المشتركة الممثلة من حيث المحتوى والخصائص الإحصائية لمجموع فقرات الاختبارات المراد معادلتها تعتبر هامة جدا عندما تكون مجموعات المفحوصين متباينة القدرة. (الصمادي،٢٠٠٧).

يذكر كل من كولن وبرينان Kolen, Brenann(1995) انه من الصعوبة بمكان الاتفاق على عدد محدد من فقرات الاختبار ذات الفقرات المشتركة، وعلى أية حال، فانه ولأغراض محددة يجب على كل باني اختبار أن يأخذ بعين الاعتبار عامل الوقت والتكلفة ومحددات السياق، بالإضافة إلى مؤشر الفاعلية عند تحديد طول اختبار الجذع المشترك.

قدم أنجوف(Angoff (1984 قاعدة تعتمد على اختيار عدد مناسب لفقرات الجذع المشترك ، وقال بأنها يجب أن لا تقل عن(٢٠) فقرة أو (٢٠%) من المجموع الكلي لعدد فقرات نماذج الاختبار (أيهما أكثر).وأشار كل من مكبرايد وويس (McBride&Weiss(1974 أن مصمم الاختبار يحتاج إلى ما بين (٤٠-٦٠) فقرة لاختبار الجذع المشترك؛ لغايات المعادلة وفقا للنظرية الكلاسيكية في القياس، وذلك وفقا للقيم النظرية للأخطاء المعيارية لتقديرات الفقرة. أما رايت (Wright (1997 فقد اعتبر انه من (١٠-٢٠) فقرة تعتبر كافية لغايات المعادلة بشرط ان تكون الفقرات ذات خصائص جيدة. وقد اقترح كل من راجو وادواردز واوبسبرغو ولورد (Raju, et al (1983) & Lord(1980 أن (٥) أو (٦) فقرات من اختبار الجذع المشترك تعد كافية لإجراء عملية المعادلة باستخدام النظرية

الحديثة في القياس، وذلك إذا تم اختيارها بحيث تكون معالم الفقرة في كلا النموذجين مقدرة بتحليل منفرد باستخدام طرق IRT (الصمادي،٢٠٠٧).

طرق معادلة الاختبارات:

تعتمد معادلة الاختبارات على طريقتين:

١- الطرق التي تعتمد على النظرية التقليدية في الاختبارات Classical Test (CTT) Theory.

٢- الطرق التي تعتمد على النظرية الحديثة في القياس التربوي (نظرية الاستجابة للفقرة) Item Response Theory (IRT).

أولاً: طرق معادلة الاختبارات باستخدام النظرية التقليدية للاختبارات:

تتلخص الطرق الرئيسية المستخدمة في معادلة الاختبارات ضمن هذه النظرية في:
معادلة الاختبارات باستخدام الوسط الحسابي، ومعادلة الاختبارات باستخدام المعادلة الخطية، والطريقة الانحدارية في معادلة الدرجات، ومعادلة الاختبارات بواسطة الرتب المئينية المتساوية(Kolen & Brennan, 1995).

(١) معادلة الاختبارات باستخدام طريقة الوسط الحسابي (Mean Equating):

تتلخص هذه الطريقة في جعل الوسط الحسابي للدرجات متساوياً في صورتي الاختبار لمجموعة المفحوصين التي تُستخدم بياناتها في إجراء عملية معادلة الاختبار؛ أي يتم إجراء تحويل رياضي يؤدي إلى تساوي متوسطي الدرجات في صورتي الاختبارين. مثلاً: إذا كان متوسط درجات المفحوصين في الصورة الأولى للاختبار هو ٧٧ درجة، ومتوسط درجاتهم في الصورة الثانية من الاختبار هي ٧٢ درجة، فذلك يعني أن الدرجة ٧٧ في الصورة الأولى يتم تحويلها رياضياً لتصبح مكافئة للدرجة ٧٢ في الصورة الثانية؛ ويكون الفرق بين الدرجتين (٧٧-٧٢ = ٥) ثابتاً خلال عملية تدريج الدرجات Calibration. فإذا حصل طالب على درجة ٧٥ في الصورة الثانية للاختبار، تكون درجته في الصورة الأولى للاختبار ٨٠ درجة. ويمكن التعبير رياضياً عن معادلة الاختبار باستخدام طريقة الوسط الحسابي كما يأتي:(الدوسري،٢٠٠٤)

$$X_1 - \overline{X}_1 = X_2 - \overline{X}_2$$

حيث (X_1 X_2) هما درجة الطالب في الصورة الأولى للاختبار ودرجته في الصورة الثانية

$\overline{X}_2$ و $\overline{X}_1$ هما المتوسط الحسابي للصورة الاولى والثانية على التوالي.

وتعني هذه المعادلة حساب درجة طالب في الصورة الأولى للاختبار من خلال درجته في الصورة الثانية لنفس الاختبار. أي أن:

$$X_1 - X_2 - 77 + 72 = X_2 - 5$$

فمثلا، إذا حصل الطالب على درجة ٧٥ في الصورة الثانية للأختبار، تكون درجته في الصورة الأولى للاختبار مكافئة للدرجة ٧٠ في الصورة الأولى. أي أن:

$$X_1 - X_2 - ٥ = 75 - 5 = 70$$

(٢) معادلة الاختبارات باستخدام طريقة المعادلة الخطية Linear Equating:

في هذه الطريقة يتم تحويل الدرجات رياضياً بحيث تتساوى المتوسطات الحسابية والانحرافات المعيارية للفقرات في صورتي الاختبار المراد معادلتهما للمجموعة المستهدفة من المفحوصين من خلال بياناتهما، وتفترض طريقة معادلة الاختبار خطياً أن الفروق في مستويات الصعوبة بين الفقرات في صورتي الاختبار المراد معادلته تكون ثابتة على امتداد سلم تدريج الدرجات. وفي كثير من الأحيان تُعتبر الفروق النسبية بين مستويات صعوبة الفقرات على أنها متغير variable على امتداد تدريج (سُلَّمْ) الدرجات. فقد تكون الصورة الأولى للاختبار أكثر صعوبة نسبياً للطلبة ذوي التحصيل المنخفض منها لدى الطلبة مرتفعي التحصيل، فالتحويلات الخطية (Linear Transformations) هنا تسمح للصعوبة النسبية للفقرات بأن تتباين على امتداد تدريج الدرجات. (الدوسري،٢٠٠٤)

ويمكن تحويل الدرجات خطيا بتحويلها من درجات خام إلى درجات معيارية زائية متساوية في صورتي الاختبار ($Z_1 = Z_2$)، كما يأتي:

$$\frac{X_1 - X_1}{S_1} = \frac{X_2 - X_2}{S_2}$$

حيث S1 و S2 هما الانحرافان المعياريان للفقرات في كل من صورتي الاختبار على التوالي. فللحصول مثلاً على درجة طالب في الصورة الأولى للاختبار من مكافئها في الصورة الثانية للاختبار، تتحول المعادلة السابقة إلى الآتي:

$$X_1 = \frac{S_1}{S_2} \quad X_2 \left[\overline{X}_1 - \frac{S_1}{S_2} \overline{X}_2 \right] = AX_2 + B$$

حيث:

$$B = \overline{X}_1 - \frac{S_1}{S_2} \overline{X}_2, \qquad A = \frac{S_1}{S_2}$$

وأن A هو ميل التحويل الخطي slope، B هو المقطع الصادي intercept.

فعلى سبيل المثال، لو افترضنا أن متوسط درجات المجموعة في الصورة الأولى للاختبار هو ٧٢، ومتوسط درجاتهم في الصورة الثانية له هو ٧٧، وأن الانحراف المعياري لفقرات كل من صورتي الاختبار هو ٩، ١٠ على التوالي فإن التحويل الخطي المطلوب للحصول على درجات معادلة لصورة من الاختبار إلى صورة أخرى هي كما يأتي:

$$X_1 = \frac{9}{10} \quad X_2 + \left[72 - \frac{9}{10}(77) \right] = 0.9X_2 + 2.7$$

فإذا حصل طالب مثلاً على الدرجة ٨٠ في الصورة الثانية للاختبار، فإن الدرجة التي تعادلها في الصورة الأولى للاختبار هي:

$$X_1 = 0.9(80) + 2.7 = 24.7$$

(٣) طريقة المعادلة باستخدام الرتب المئينية المتساوية (Equipercentile Equating):

في هذه الطريقة نجعل التوزيع الإحصائي للدرجات في الصورة الثانية للاختبار مساوياً للتوزيع الإحصائي للدرجات في الصورة الأولى لمجموعة المفحوصين. لذلك يتم تحويل درجات الصورة الثانية للاختبار رياضياً باستخدام الرتب المئينية المتساوية، بحيث يكون له نفس الوسط الحسابي والانحراف المعياري، ونفس شكل التوزيع الإحصائي تقريباً (الالتواء والتفلطح، .. إلخ) كنظيره في الصورة الأولى للاختبار، وهذه التحويلات الرياضية في معادلة الاختبار بهذه الطريقة يمكن استخدامها في التصاميم المختلفة المستخدمة في معادلة الاختبارات. ومن هنا، فإن الرتب المئينية المتساوية لدرجات المفحوصين في صورتي الاختبار تدل على تساوي مستوى الأداء في كلتا الصورتين. وهذا هو أساس معادلة الاختبار بهذه الطريقة. أما الجانب الإجرائي في معادلة الاختبار بهذه الطريقة، فتتم بواسطة رسم بياني لدرجات الطلبة في صورتي الاختبار معاً يبين التوزيع التكراري النسبي المتجمع (الرُتب المئينية مقسومة على ١٠٠). وبذلك تكون درجات الطالبة في أي من الصورتين متكافئة إذا تساوت رتبها المئينية. :(الدوسري،٢٠٠٤)

مقارنة بين الطرق التقليدية لمعادلة الاختبارات:

تتلخص الفروق بين طرق معادلة الاختبار التقليدية في النقاط الآتية:

- يمكننا القول بشكل عام أنه إذا كانت صورتا الاختبار المراد معادلتهما لهما نفس مقدار الانحراف المعياري، فإن معادلة الاختبار بطريقة الوسط الحسابي وبطريقة المعادلة الخطية تؤديان إلى نفس النتائج. أما إذا كان شكل التوزيع الإحصائي للدرجات في الصورتين هو نفسه (نفس الالتواء والتفلطح، إلخ..)، فإن طريقة معادلة الخطية للاختبار وطريقة الرُتب المئينية المتساوية تعطيان نفس النتائج.
- تتطلب معادلة الاختبار بطريقة الرُتب المئينية المتساوية عينة كبيرة من المفحوصين، لأنها أكثر تعقيداً من الناحية الرياضية من طريقة الوسط الحسابي وطريقة المعادلة الخطية. وينطبق ذلك بشكل خاص على تصميم المجموعات غير المتكافئة الذي يتضمن فقرات اختبارية مشتركة بين صورتي الاختبار.

- معادلة الاختبار خطيا هي عملية تحليلية بحتة، بخلاف طريقة الرُتب المئينية المتساوية.
- لا يمكن للتحويل الرياضي بطريقتي المعادلة الخطية و الرُتب المئينية المتساوية أن يُنْتِجَ معادلة اختبار صحيحة وتامة ما لم يتم استيفاء شرط العدالة (المساواة) Equity. (Kolen & Brennan, 1995)

ثانيا: معادلة الاختبارات باستخدام نظرية استجابة الفقرة:

إن معادلة الاختبارات المعتمدة على الدرجات الخام Raw Scores ضمن النظرية التقليدية للاختبارات قد لاتكون مرغوبة؛ وذلك بسبب إخفاقها في تحقيق جميع شروط معادلة الاختبارات التي أشرنا إليها سابقا، وهي (العدالة والمساواة، واللاتباين) (Hambleton & Swaminathan (1985) لذلك، فإن معادلة الاختبار عن طريق نظرية الاستجابة للفقرة (النظرية الحديثة في القياس التربوي والنفسي) Item Response Theory تحل الكثير من المشكلات التي عجزت عن حلها النظرية التقليدية في الاختبارات؛ بشرط أن يكون نموذج IRT Model المستخدم في النظرية الحديثة هذه مطابقاً للبيانات المُعدة لعملية المعادلة.

ويشير بيكر والقرني (Baker & Alkarni (1991, p 147. إلى أنه من الإسهامات الكبيرة للنظرية الحديثة في القياس التربوي والنفسي، وخصوصا في الممارسة التربوية، قدرة هذه النظرية على وضع عدة اختبارات ومجموعات من المفحوصين على تدريج مشترك Common Scale في عملية القياس؛ وإمكانية استخدامها في المعادلة الأفقية والرأسية للاختبار.

تقوم نظرية الاستجابة للفقرة على مجموعة من الافتراضات، حيث تفترض هذه النظرية أن أداء المفحوصين في الاختبار يمكن تفسيره عن طريق السمة أو السمات الكامنة Latent Traits المراد قياسها، والتي لا يمكن قياسها بصورة مباشرة. إذ يتم استخدام الدرجات التي تم تقديرها للمفحوص في تلك السمة في التنبؤ بادائه في اختبار ما أو في فقرة من الاختبار؛ لأن العلاقة الحقيقية بين الدرجات المشاهدة (الخام) للمفحوص والسمة المراد قياسها لا يمكن الحصول عليها بطريقة مباشرة. ومن هنا، فان نظرية الاستجابة للفقرة تصف هذه العلاقة بواسطة دالة تعتمد على مجموعة من الافتراضات،هي: أحادية البُعْدْ (أحادية السمة) Unidimensionaity حيث يقيس الاختبار سمة واحدة فقط؛ والاستقلال الموضعي Local

Independence وهو استقلال أداء المفحوص على فقرة اختبار عن أدائه على فقرة أخرى. وكذلك جعل السمات الأخرى التي تؤثر على أداء المفحوص ثابتة ومتسقة Consistent. أما الافتراض الأخير فهو افتراض اللاتباين Invariance والذي يعني أن معالم (معلمات) Parameters الفقرة (الصعوبة، والتمييز، والتخمين) لا تعتمد على التوزيع الإحصائي للسمة المراد قياسها؛ وأن المعالم التي تصف أداء المفحوصين لا تعتمد على فقرات الاختبار (Hambelton,et al, 1991).

خطوات معادلة الاختبارات باستخدام نظرية استجابة الفقرة:

وضح كل من هامبلتون وسواميناثان (Hambleton & Swaminthan (1985، وكذلك كولن وبرينان (Kolen & Brennan(1995، الخطوات الضرورية لمعادلة الاختبارات بواسطة النظرية الحديثة في القياس التربوي والنفسي وهي كالآتي:

- اختيار التصميم المناسب لمعادلة الاختبار مع الأخذ بعين الاعتبار خصائص مجموعة المفحوصين وطبيعة الاختبارات المراد معادلتها.
- اختيار النموذج المناسب الذي يطابق التصميم المناسب والاختبار المناسب (نموذج راش أو غيره من نماذج هذه النظرية).
- بناء تدريج مشترك يربط العلاقة بين السمة المراد قياسها ومَعْلم الفقرة Item Parameter.
- اختيار التدريج المناسب لوضع درجات الاختبار. أي هل تُكْتَبُ الدرجات كدرجات خام (درجات ملاحظة)، أو على صورة درجات قدرة Ability Scores، أو على صورة درجات حقيقية مقدرة Score Estimated True، ويمكن أن تقوم بهذه المهمة المعقدة رياضياً بعض البرامج الحاسوبية، مثل برنامج BILOG، وبرنامج MULTILOG وبرنامج LOGIST، وغيرها. ومن المتعذر عملياً إلى حدٍ كبير القيام بهذه العملية يدوياً، وخاصة في هذه النظرية.

ويوضح كوك وايجنور (Cook & Eignor (1991 الآليات الأساسية لعملية معادلة الاختبارات في هذه النظرية، والتي يمكن إيجازها في الآتي:

- اختيار التصميم المناسب: هناك ثلاثة تصاميم أساسية تستخدمها هذه النظرية لمعادلة الاختبارات؛ وهي تصميم المجموعة المفردة، وتصميم المجموعات العشوائية، والتصميم ذو الاختبار المشترك. ويعتمد حجم العينة الملائم، لإجراء معادلة الاختبار بشكل صحيح، على

العدد الملائم للمفحوصين، للحصول على تقديرات مستقرة للمعلمات المستخدمة في النموذج المختار؛وذلك لوضع الدرجات ومعلمات القدرة (السمة) على تدريج واحد. وفي عملية المعايرة هذه نحتاج إلى عينة تصل إلى ٣٠٠٠ مفحوص. وإذا كانت هناك فقرات مشتركة بين صورتي الاختبار على هيئة اختبار مشترك Anchor Test، فيجب أن تعكس هذه الفقرات المحتوى والخصائص الإحصائية لصورتي الاختبار؛ وأن لا تقل نسبة فقرات الاختبار المشترك عن ٢٠% من الطول الاحتمالي للاختبار. كما أن اختيار أي من التصميمات المشار إليها أعلاه يرتبط بنوعية البرنامج الحاسوبي المستخدم لمعادلة الاختبار.

- وضع تقديرات المعالم (المعلمات) على تدريج مشترك: لو افترضنا أن مجموعتين من المفحوصين طُبقت عليهما نفس المجموعة من فقرات الاختبار، وتم تقدير معلمات الفقرة (الصعوبة والتمييز) لكل مجموعة على حده، وأن النموذج الرياضي المستخدم هو النموذج ثنائي المعلم؛ حيث أن منحنيات خاصية الفقرة Item Characteristic Curves (ICC) مستقلة عن المجموعتين المستخدمتين لرسم هذه المنحنيات. فانه يمكن أن نستنتج من ذلك أن تقديرات معلمات الفقرات متطابقة في كل من المجموعتين على حده. ولكن الواقع ليس كذلك! إذ يجب إجراء تحويل رياضي معين يوحد التدريج في عملية المعايرة. ويعتمد ذلك على نوع البرنامج الحاسوبي المستخدم في إجراء عملية التحويل الرياضي للحصول على نقطة أصل ووحدة قياس للسمة ولمستوى الصعوبة. ويكون متوسط درجات القدرة (السمة) هو الصفر وانحرافها المعياري هو الواحد الصحيح (برنامج LOGIST يمكن أن يحقق ذلك). ويجب ملاحظة أن عملية التحويل الرياضي التي تضع الدرجات وتقديرات معلمات السمة على نفس التدريج تستهدف الحصول على معلمَتي المَيْلْ Slop والمقطع الصادي Intercept، بحيث يكون الوسط الحسابي والانحراف المعياري لتوزيع مستويات صعوبة الفقرات، التي تم تقديرها في عملية المعايرة الثانية للدرجات، مساويين لنظيريهما اللذين تم تقديرهما في المعايرة الثانية للدرجات، مساويان لنظيريهما اللذين تم تقديرهما في المعايرة الأولى للدرجات.

- معادلة درجات الاختبار: تُعتبر عملية معادلة الاختبار منتهية إذا تم تقدير المعلمات لفقرات كل من صورتي الاختبار المستهدف، وتم وضعها على تدريج مشترك، وتم كذلك الحصول على تقدير للسمة المراد قياسها لدى المفحوص؛ بحيث تكون هي نفسها في أي من صورتي الاختبار. ويؤخذ خطأ القياس بعين الاعتبار هنا. وبناء على ذلك، يكون التعبير عن الدرجات الخام بما يكافئها من درجات السمة. أما إذا أخفق البرنامج الحاسوبي المستخدم في استخراجها، يلجأ المختصون إلى ترجمة أو تحويل أي درجة من درجات القدرة إلى الدرجة

الحقيقية المقدرة المناظرة لها في صورتي الاختبار؛ واعتبارها الدرجة التي تمت معادلتها في الاختبار. أما الصورة الرياضية للدوال التي تربط بين درجات القدرة وتقديرات الدرجات الحقيقية فهي كالتالي:

$$\hat{T}_x = \sum_{1=1}^{n} \hat{P}_i(\theta)$$

$$\hat{T}_y = \sum_{j=1}^{n1} \hat{P}_j(\theta)$$

$\hat{T}_x$ = الدرجة الحقيقية المقدرة للصورة الأولى من الاختبار.

$\hat{T}_y$= الدرجة الحقيقية المقدرة للصورة الثانية من الاختبار.

$\hat{P}_j(\theta)$= الدالة المقدّرة لاستجابة الفقرة في الفقرات للصورة الأولى للاختبار.

$\hat{P}_j(\theta)$= الدالة المقدّرة لاستجابة الفقرة للصورة الثانية في الاختبار. علماً بأن التحويل الرياضي للدرجات في كلٍ من صورتي الاختبار، يكون مستقلاً عن مجموعة المفحوصين التي تم الحصول على بيانات معادلة الاختبار منها لإجراء هذه التحويل. ويجب أن نلاحظ هنا أنه إذا كانت الصورة القديمة للاختبار المراد معادلته أكثر صعوبة من الصورة الجديدة في بعض المستويات، فإنها تُعطي تقديراً منخفضاً للدرجة الحقيقية المطلوب الوصول إليها عن طريق تقدير درجة السمة.(الدوسري، ٢٠٠٤)

الفوائد العملية والتطبيقية لمعادلة الاختبارات باستخدام النظرية الحديثة في القياس

تشير نتائج الأبحاث التي أجريت على معادلة الاختبار بواسطة هذه النظرية، إلى أن معادلة الاختبار بهذه النظرية لها فوائد جمة من الناحيتين العملية والتطبيقية يمكن إيجازها في الآتي:

١. معادلة الاختبار بهذه النظرية هي الأفضل عندما تكون فقرات الاختبارات المختلفة في مستويات الصعوبة، مطبقة على مفحوصين من مجموعات غير عشوائية، ومختلفة في مستويات السمة المراد قياسها.

٢. معادلة الاختبار بهذه النظرية، وبسبب خصائصها المشروحة سلفاً، تؤدي إلى تحويلات رياضية للدرجات تكون مستقلة عن مجموعة أو مجموعات المفحوصين التي طُبقت عليها الاختبارات.

٣. المعادلة بنظرية IRT هي أفضل من نظريتها في النظرية التقليدية للقياس التربوي، وخاصة في النهايات العليا لتدريج الدرجات، حيث غالباً ما تتم عملية اتخاذ القرارات الهامة، بحيث يمكن معادلة الدرجات الخام مع كل قيم درجات السمة.

٤. المعادلة بهذه النظرية تسمح باستخدام الصورة القديمة للاختبار لمعادلة درجاتها بالصورة الجديدة للاختبار متى ما تم وضع كل من الدرجات في كلا الاختبارين، وكذلك تقديرات معلمات القدرة، على تدريج واحد.

٥. معادلة الاختبار بهذه النظرية تسمح بمعادلة الاختبار على مستوى الفقرة الواحدة، وذلك قبل الشروع في تطبيق الاختبار في صوره المتعددة، وقبل تطبيق صورتيه أيضاً. ويتم ذلك إذا توافرت البيانات القبلية على مستوى الفقرة الواحدة، مع إمكانية معايرتها . ثم يتم وضع تقديرات المعلمات على تدريج مشترك، وهذه الخاصية لا يمكن الحصول عليها من خلال معادلة الاختبار بواسطة النظرية التقليدية في الاختبارات. (Kolen, 1988, P. 149).

يمكن التعبير عن العلاقة الرياضية بين درجة المفحوص (الدرجة الخام) على الفقرة ودرجته على سُلّمْ السّمة المراد قياسها بعدة نماذج رياضية في نظرية استجابة الفقرة وهي: نموذج راش Rash أحادي المَعْلمة (نسبة إلى العالم الرياضي الدنمركي جورج راش)، والنموذج ثنائي المَعْلمةTwo-Parameter Model، والنموذج ثلاثي المَعْلمة Three-Parameter Model ؛ ويعبر عن هذه النماذج رياضيا كما ياتي:

نموذج راش أحادي المَعْلَمْ

$$P(\theta) = \frac{1}{1 + e^{D(\theta - b)}}$$

النموذج ثنائي المَعْلَمْ

$$P(\theta) = \frac{1}{1 + e^{Da(\theta - b)}}$$

النموذج ثلاثي المَعْلَمْ

$$P(\theta) = \frac{1}{1 + e^{Da(\theta - b)}}$$

حيث:

$P(\theta)$ = احتمال السمة المراد قياسها.

a = معامل تمييز الفقرة (قيمته ثابتة في نموذج راش).

b = معامل صعوبة الفقرة.

c = مَعْلَمْ التخمين للفقرة.

D = مَعْلَمْ تدريج Scale Parameter يجعل الدالة اللوجستية Logistic Function قريبة إلى الحد الأقصى من دالة المنحنى الطبيعي التراكمي Normal Ogive Function وقيمة هذا المَعْلَمْ تساوي 1.7

e = الثابت الرياضي الذي يحول الدالة التي تربط بين الدرجة الخام والسمة المراد قياسها، من دالة ما لانهاية إلى دالة احتمالية تحصر العلاقة بين أدء الطالب على الفقرة وصعوبة الفقرة بين الصفر والواحد الصحيح. وقيمة هذا الثابت الرياضي تساوي 2.72

النماذج المتعددة الاستجابة في نظرية استجابة الفقرة:

إن نماذج نظرية استجابة الفقرة والتي تمثل العلاقة ما بين قدرة الافراد واستجاباتهم لا يمكن تطبيقها على الاستجابات التي لا تحتمل امكانية الصواب أو الخطأ ، وبمعنى آخر الاستجابات التي يمكن أن تكون وفق سلم للاجابات.

ان المقاييس المستخدمة في قياس الشخصية والاتجاهات والتي تتضمن فقرات يكون الاستجابات عليها بناء على سلم مقسم إلى فئات، اكثر صدقا من الاستجابات التي تعتمد على وجود بديلين وهما: الصح والخطأ، وتعطي معلومات اكثر حول المفحوصين. ولمثل هذا النوع من الاستجابات ظهرت هناك حاجة إلى وجود نماذج رياضية توضح العلاقة غير الخطية ما بين قدرة الافراد، وما بين احتمالية الاجابة في فئة معينة من فئات الاجابات (Embreston, 2000).

لذلك سيتم الحديث في هذه الدراسة عن خمسة نماذج رياضية تعاملت مع الاستجابات المتعددة او وفق سلم اجابات محدد، وهذه النماذج هي:

- Graded response model (GRM).
- Modified Graded response model (M-GRM)
- Partial Credited model (PCM).
- The Generalized Partial Credit Model .(G-PCM)
- Rating Scale Model (RSM).

يعتبر النموذج الاول والثاني نماذج غير مباشرة من نماذج نظرية استجابة الفقرة؛ وذلك لانهما يتطلبان وجود خطوتين لتحديد احتمالية الاستجابة المشروطة للفرد ضمن فئة معينة من الاجابة، إلا أن النماذج الاخرى يعتبر من النماذج المباشرة لنظرية استجابة الفقرة؛ وذلك لكونهما يحتاجان إلى خطوة أو معادلة واحدة لوصف العلاقة ما بين قدرة المفحوص واحتمالية اجابة الفرد ضمن فئة معينة من الاستجابات.

أولا:نموذج الاستجابات المتدرجة.Graded Response Model (GRM)

وضع هذا النموذج من قبل سيميجما Samejima (1969) ليتم استخدامه مع الاستجابات التي تتم وفق فئات مرتبة مثل استخدام سلم ليكرت للتقدير، ويعتبر هذا النموذج جزءا من النموذج الثنائي لنظرية استجابة الفقرة والذي يصنف ضمن فئة النماذج التي تعتبر نماذج غير مباشرة من نماذج نظرية استجابة الفقرة، وذلك لأن عملية حساب احتمالية اجابة المفحوصين ضمن فئة معينة تتطلب عملية مكونة من خطوتين.

ليلائم هذا النموذج مقياسا معينا، فإن الفقرات لا تتطلب تساوي عدد فئات الاجابات، بمعنى أنه لا يوجد تعقيدات في عملية تقدير معالم الفقرات ناتج عن استخدام مقياس معين لفقرات لها نماذج مختلفة من الاجابات، ولكن ذلك لا ينطبق على نموذج Rating Scale Model والذي سيتم مناقشته لاحقا. (Embreston, 2000)

وفقا لهذا النموذج (GRM) فإن كل فقرة والتي يرمز لها بالرمز (i) يمكن وصفها من خلال معلم تمييز واحد (αi) يقع ما بين معالم الفئات أو الفواصل لهذه الفئات Threshold والذي يرمز له بالرمز (βi) ، وهناك مرحلتان في عملية حساب احتمالية الاجابة لكل فئة في هذا النموذج، ولفهم هذه المراحل، نفترض أن لدينا فقرة اختبار مكونة من خمسة بدائل (K=5) والتي يحصل من خلالها المفحوص على علامات ما بين (0-4) (4.....٠ =α) وبما أن لدينا (5) بدائل للإجابة فإن عدد الفواصل ستكون (4) وهي الحدود ما بين الفئات (mj = 4) و (j =1,2,3,4) حيث أن أحد الأهداف لاختبار ملائمة هذا النموذج هو تحديد موقع هذه الحدود على متصل القدرة.

إن الخطوة الأولى في هذا النموذج هي تقدير احتمالية الاجابة للفقرة والتي تعتمد على عملية تقدير منحنيات الفئة (mi) لكل فقرة وباستخدام المعادلة التالية: (Embreston, 2000)

$$P_{ix}(\theta)=\frac{\exp\left[\alpha_i(\theta-\beta_{ij})\right]}{1+\exp\left[\alpha_i(\theta-\beta_{ij})\right]}$$

Where

$X = j = 1, \ldots, m_j.$

حيث:

P (θ) = احتمال السمة المراد قياسها.

a = معامل تمييز الفقرة

ßij = معامل صعوبةالفقرة.

Θ = معلمة القدرة

عند التدقيق في المعادلة السابقة، نلاحظ أنها تمثل النموذج ثنائي المعلمة من نماذج نظرية استجابة الفقرة، حيث أن لكل منحنى من المنحنيات التي تم تبيانها في المعادلة السابقة تمثل احتمالية إجابة المفحوص أو العلامة الخام (X) والتي تقع في أو أعلى من الحد الفاصل للفئة (j = 1 ...mi) .

هذا وتسمى المنحنيات التي تمثل احتمالية اجابة المفحوص ضمن كل فئة (ICC)، ووفقا لهذا النموذج (GRM) ، يجب تقدير منحنى واحد فقط لكل حد أو فاصل ما بين الفئات، وبذلك فإن الإجابات المكونة من (٥) فئات له (٤) معالم للصعوبة (βij) ،ومعلم واحد للتمييز (αi)، ويتم تفسير معلمة الصعوبة (βij) وفقا لهذا النموذج كما يأتي:
تمثل هذه القيم مستوى القدرة المطلوب للإجابة اعلى من الفاصل والذي يرمز له بالرمز (j) باحتمالية (٥٠%)، وبشكل عام فإن ما يحدث في هذا النموذج هو أن الفقرات يتم التعامل معها على أنها سلسلة من الاستجابات (mi = k -1) ويتم تقدير النموذج الثنائي لكل سلسلة مع التركيز على أن معلم التمييز لمنحنيات الفقرات هو ثابت، وعندما يتم تقدير خصائص هذه المنحنيات (ICC) فإن عملية تقدير احتمالية الاستجابة الحقيقية لكل فئة (x = 0 ...4) يتم حسابها من خلال المعادلة التالية: (Embreston, 2000)

$$P_{ix}(\theta) = P_{ix}^*(\theta) - P_{i(x+1)}^*(\theta)$$

حيث:

$P_{ix}(\theta)$:هي احتمالية حصول الفرد الذي قدرته (θ) على القيمة ($_x$) للفقرة (i)

ومن خلال التعريف فإن احتمالية الاجابة ضمن الفئة،أو أعلى من الفئة الأدنى هي (P*io(θ)=1) واحتمالية الاجابة أقل من أعلى فئة هي (P*i5(θ)=0) ، وتكون احتمالية الاجابة في كل فئة من الفئات الخمسة كما ياتي:

$$P_{i0}(\theta) = 1.0 - P_i^*(\theta)$$

$$P_{i1}(\theta) = P_{i1}^*(\theta) - P_{i(2)}^*(\theta)$$

$$P_{i2}(\theta) = P_{i2}^*(\theta) - P_{i(3)}^*(\theta)$$

$$P_{i3}(\theta) = P_{i3}^*(\theta) - P_{i(4)}^*(\theta)$$

$$P_{i4}(\theta) = P_{i4}^*(\theta) - 0$$

وتسمى هذه المنحنيات للاجابات (CRC) وهي تمثل احتمالية اجابة المفحوص في فئة معينة والتي تعتمد على مستوى القدرة.

إن معالم الفقرات في (GRM) تبين شكل منحنيات فئات الإجابة، ومنحنى الخصائص (ICC)، وبشكل عام فإنه كلما ارتفع أو زاد الميل (aij) كان منحنى خصائص الفقرة أكثر حدة، كما أن معالم الفواصل بين الفئات (Bij) تحدد مكان منحنيات خصائص الفقرات (ICC)، لذلك فإن استخدام برنامج مثل MULTILOG يعمل على توفير بيانات عن كل من العلامة المشاهدة ونسبة الاستجابات في كل فئة، اضافة إلى القيم التنبؤية لهذه المعالم (بمعنى التقدير من خلال معالم الفقرات وتوزيع مستويات القدرة) حيث أن الاختلاف في هذه القيم يشير إلى قدرة هذه النماذج على التنبؤ بالاستجابات الحقيقية للفقرات .(Embreston, 2000)

ثانيا:نموذج الاستجابة المعدل المتدرج

MODIFIED GRADED RESPONSE MODEL

قام موركي(1990) Murki بتطوير نموذج معدل لـــ (GRM) لاستخدامه في تحليل الاستبيانات التي لها طابع الاستجابات المتعددة (Rating Scale) مثل الاستبيانات الخاصة بالاتجاهات والتي يكون لنفس الفقرة فئات للإجابة والمسمى بـــ (M-GRM).

ويعتبر هذا النموذج حالة خاصة من نموذج (GRM)، فهو يتيح المجال لميل الفقرات أن يتفاوت في القيم ما بين الفقرات، كما أنه يعتبر نموذجا غير مباشر من نماذج نظرية استجابة الفقرة (لأنه يتطلب القيام بخطوتين في حساب احتمالية الاجابة). ووفقا لهذا النموذج فإن الفئات ما بين معالم الحدود الفاصلة (Bij) أي الصعوبة، قد تم تقسيمها إلى قسمين وهما: (Embreston, 2000)

١- موقع معلم صعوبة الفقرة (bi) لكل فقرة.

٢- مجموعة من معالم الفواصل ما بين الفئات (ci) على طول المتصل أي Bij=bi+ci

كما أن منحنى خصائص الفقرة (ICC) لهذا النموذج يمكن كتابته بالشكل الآتي: (Embreston, 2000)

$$P_{ix}(\theta) = \frac{\exp\left[\alpha_i(\theta - (b_i - c_j))\right]}{1 + \exp\left[\theta - (b_i - c_j))\right]}$$

حيث: (bi) : موقع صعوبة الفاصل.

Ɵ: معلمة القدرة.

Ci: هي صعوبة الحدود الفاصلة على طول المتصل

وبالرجوع الى معادلة هذا النموذج، نلاحظ انها تبين احتمالية الاجابات في فئة معينة او اعلى منها كما هو في نموذج GRM، كما ان معلم التمييز في هذا النموذج هو شبيه بمعلم التمييز في النموذج السابق، والذي يبين سرعة التغيير في علامة الفقرات المتوقعة نتيجة تغير عامل او مستويات القدرة، الا ان الاختلاف بين هذا النموذج والنموذج السابق هو ان هنالك قيمة واحدة لصعوبة الفاصل او الحدود الفاصلة (Bij) يتم تقييمها لكل فقرة في نموذج GRM،

بينما في النموذج المعدل، يوجد هنالك تقدير لصعوبة الحدود الفاصلة (Cj) على طول المتصل، وهنالك قيمة للصعوبة عند كل فقرة من فقرات bj، حيث ان الفقرة تعمل على توضيح حركة او تغيير معالم صعوبة الفاصل للفقرات على طول متصل القدرة والتي تسمى بالصعوبة، ومن المتوقع ان يحصل بعض المفحوصين على اجابات صحيحة في الفئات التي يوجد لها قيم ايجابية عالية لـــ(bj) اي ان مستوى صعوبة الفقرة تكون عالية والعكس صحيح.

ان سبب وجود هذا النموذج هو الحاجة الى وصف الاجابات على سلم ليكرت والمستخدم في مقاييس الاتجاهات، حيث يتيح المجال للفصل ما بين تقدير معالم الفقرة ومعالم الحدود الفاصلة ما بين فئات الاجابات.

ثالثا: نموذج التقدير الجزئي .THE PARTIAL CREDIT MODEL(PCM)

تم تطوير هذا النموذج للفقرات التي تتطلب الاجابة عليها مجموعة من الخطوات حيث يتم اعطاء علامة لكل خطوة منها، وغالبا ما يتم استخدامه في الاختبارات التحصيلية مثل: اختبارات الرياضيات والتي يتم الحصول على جزء من العلامة حسب خطوات الحل، كما ان هذا النموذج ملائم في تحليل اختبارات الميول والاتجاهات والتي يعبر المفحوصين فيها عن معتقداتهم ضمن ترتيب معين. وعلى العكس من النموذجين السابقين فان هذا النموذج(PCM) هو نموذج مباشر لنظرية استجابة الفقرة؛ وهذا يعني ان احتمالية الاجابة في فئة معينة يمكن كتابتها بشكل مباشر على انها (e) مقسومة على مجموع الـــ e (Embreston, 2000) .

ان هذا النموذج يمكن اعتباره على انه فرع من النموذج الاحادي المعلمة، وله نفس خصائص نموذج راش من حيث: الفصل ما بين معالم الافراد ومعالم الفقرات، ويعبر عن هذا النموذج رياضيا كما يلي :- (Embreston, 2000)

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^{x}(\theta - \delta_{ij})\right]}{\sum_{r=0}^{m_i}\left[\exp\left[\sum_{j=0}^{r}(\theta - \delta_{ij})\right]\right]}$$

Where

$$\sum_{j=0}^{x}(\theta - \delta_{ij}) = 0$$

ان (δ_{ij}) تشير الى صعوبة الفقرة المرتبطة بالفئة، فكلما كانت هذه القيمة اعلى كان ذلك مؤشرا على زيادة صعوبة الخطوة مقارنة مع الخطوات الاخرى لكل فقرة. هذا ويمكن تفسير (δ_{ij}) بشكل مباشر على انها النقطة على متصل القدرة والتي عندها يتقاطع اثنتين من منحنيات الاجابة حسب الفئة، ولهذا السبب فان معلم (δ_{ij}) يشير في هذا النموذج الى تقاطع الفئات CATEGORY INTERSECTION ، ووفقا لهذا النموذج فانه يفترض ان يكون لكل فقرة نفس الميل، وبالرجوع الى معادلة هذا النموذج التي تشير الى ان احتمالية اجابة المفحوص ضمن فئة (X) على الخطوة (Mi) هي نتيجة الاختلاف ما بين مستوى القدرة للمفحوصين وتقاطع والانتقال من فئة الى اخرى، فان معلم (δ_{ij}) يمكن اعتباره على انه صعوبة الخطوة (mi) وقاطعها للفقرة. ولفهم هذا النموذج PCM نفترض الرسم التالي والذي يمثل ٤ فئات لفقرة تقيس الاتجاهات :

0-----------------------1---------------------------2----------------------3

Not at all   a little   moderately   agree

Step 1   step 2   step 3

نلاحظ انه يمكن للمفحوص من خلال هذه الفقرة ان يكمل ثلاث خطوات للاجابة على الفئة العليا؛ اي على المفحوصين ان يقرروا الاجابة في اي فئات الاجابة يجب ان تكون، وهنا تجدر الاشارة الى ان (δ_{ij}) يمثل معلم صعوبة الخطوة، ففي بعض الفقرات نجد ان هناك بعض الخطوات اصعب من الخطوات الاخرى، وبعضها سهل، كما ان معلم (δ_{ij}) يشير الى الموقع على متصل القدرة والذي يشير الى مكان وقوع فئة الاستجابة او تقاطع منحنى استجابة الفئة. (Embreston, 2000) .

وهناك طريقة اخرى لوصف العلاقة ما بين مستويات القدرة للمفحوصين والاستجابة على الفقرات، وهي استخدام معالم الفقرات في هذه النموذج عن طريق رسم العلامات المتوقعة (واحيانا يشار اليها بالعلامات الحقيقية) للفقرة لكل مستوى من مستويات القدرة وذلك باستخدام المعادلة التالية : (Embreston, 2000)

$$E(X)=\sum_{j=0}^{m_i} xP_x(\theta)$$

اذا ما تم جمع هذه المنحنيات التي تمثل العلامات الخام للمفحوصين في مستوى معين من مستويات القدرة لكل الفقرات، فان الناتج يكون منحنى يمثل العلامة الخام للمفحوصين في مستوى معين من مستويات القدرة.

واخيرا فان هذا النموذج هو عبارة عن نموذج راش، لان كل الفقرات لها نفس الميل (١) مما يؤدي الى تشابه المنحنيات لهذه الفقرة وهي منحنيات تراكمية.

رابعا :نموذج التقدير الجزئي المعمم THE GENERALISED PARTIL CREDIT MODEL

لقد قام (1992), MURAKI بتطوير نموذج عام لنموذج PCM والذي يسمح للفقرات ضمن مقياس معين بالتغير في معلم الميل (التمييز) وقد سمي ب (G-PCM). ويمكن تقدير المعالم في هذا النموذج باستخدام برنامج PARSCALE ، ويمكن التعبير رياضيا عن معادلة هذا النموذج كما يلي: (Embreston, 2000)

$$P_{ix}(\theta)=\frac{\exp\left[\sum_{j=0}^{x}\alpha_i(\theta-\delta_{ij})\right]}{\sum_{j=0}^{M}\left[\exp\left[\sum_{j=0}^{r}\alpha_i(\theta-\delta_{ij})\right]\right]}$$

Where

$$\sum_{j=0}^{x}\alpha_i(\theta-\delta_{ij})\equiv 0$$

حيث:

(δ_{ij}) تمثل النقاط على متصل القدرة والتي تكون الاجابة فيها على فئة معينة اعلى من سابقتها في هذا النموذج(معامل صعوبة الفئة).

ai: تمييز الفقرة.

Θ: هي قدرة المفحوص.

تتداخل معالم الفئات (δ_{ij})التي تمثل النقاط على متصل القدرة والتي تكون الاجابة فيها على فئة معينة اعلى من سابقتها في هذا النموذج، اي ان المفحوص قد قام بحل الخطوات السابقة، الا ان معلم الميل في هذا النموذج (α_i) لا يتم تفسيره كما في نموذج PMC؛ والسبب في ذلك هو ان النماذج التي لايكون فيها الاجابة (0,1) ، يكون تمييز الفقرة على علاقة مباشرة مع مدى انتشار فئات صعوبة الفواصل او تقاطعات الفئات، فالميل في هذا النموذج يمثل درجة اختلاف فئات الاستجابات ما بين الفقرات باختلاف مستويات القدرة، فكلما كان الميل قليلا (اقل من ١) فان منحنيات الاجابة الفئات CRC تكون اكثر تفلطحا، وكلما كان الميل اكثر من (١) كان عند اقصى نقطة (قمة) PEAK في المنحنى.(Embreston, 2000)

خامسا :نموذج تقدير المقياس RATING SCALE MODEL

تتعدد طرق حساب احتمالية الاجابة وفقا لهذا النموذج، وعلى الرغم من ان هذا النموذج يمكن اشتقاقه من نموذج(PCM) الا ان هناك فرقا اساسيا بين النموذجين، فالفقرات لها نفس نمط الاجابة كما هو في PCM ، الا انه وفقا لهذا النموذج فان كل فقرة يتم وصفها بشكل فردي(اي معلم واحد يسمى لمبدا λ) والتي تمثل صعوبة او سهولة الفقرة، كما ان (J) والتي تمثل J=K-1 (وهي الفواصل بين الفئات) والذي يرمز له بالرمز (δ) (DELTA) تكون متساوية ما بين الفقرات. وكل فقرة يتم وصفها من خلال قيمة معلم (λ) . وكما هو في النموذج PCM فان معلم التمييز هو متساوي في جميع الفقرات والعلامة الخام هي تقدير لمستويات قدرة الافراد.

ان الاختلاف ما بين هذا النموذج وPCM يتلخص في كون نموذج PCM لا يفترض وجود علاقة ما بين صعوبة الخطوات؛ بمعنى ان الخطوة رقم (٢) مثلا ليس بالضرورة ان تكون اصعب من الخطوة (١)؛ اي ان مدى انتشار التقاطعات ما بين معالم الفقرات يمكن ان

يختلف من فقرة الى اخرى. ففي هذا النموذج فان صعوبة الخطوة يرجع الى كل من قيم (λ ,δ) بحيث (λ) هي موقع الفقرة على سلم القدرة. و(δ) هي معالم الفئات، ويمكن التعبير عن هذا النموذج بالمعادلة التالية : (Embreston, 2000)

$$P_x(\theta) = \frac{\exp\left\{\sum_{j=0}^{x}[\theta(\lambda_i + \delta_i)]\right\}}{\sum_{x=0}^{M}\left[\exp\left\{\sum_{j=0}^{x}[\theta(\lambda_i + \delta_i)]\right\}\right]}$$

Where

$$\sum_{j=0}^{x}[\theta(\lambda_i + \delta_i)] = 0$$

حيث:

بحيث (λ) هي موقع الفقرة على سلم القدرة. و(δ) هي معالم الفئات

الا ان بعض الباحثين اشاروا الى هذا النموذج بالمعادلة التالية: (Embreston, 2000)

$$P_x(\theta) = \frac{\exp[\psi_x + x(\theta - \lambda_i)]}{\sum_{x=0}^{M}\exp[\psi_x + x(\theta - \lambda_i)]}$$

Where

$$\psi_z = -\sum_{j=0}^{x}\delta_i$$

$$\psi_0 = \psi_m = 0$$

ويفترض هذا النموذج ان هناك مجموعة من النقاط RATING POINTS التي يمكن استخدامها لكل الفقرات، فاذا وجدت نماذج مختلفة للاجابات، فان هذا النموذج لا يعتبر مناسبا.

ولكن وبوجود برنامج RUMM فانه يمكن تقدير المعالم لمجموعة من الفقرات التي لها نفس عدد فئات الاجابة (الخطوات) ضمن وحدة واحدة، حيث ان اداة القياس يمكن ان تتضمن اثنين او ثلاث مجموعات من RMS وبنماذج اجابات مختلفة، الا ان المشكلة في تحليل هذا النمط من الفقرات يكمن في محاولة معادلة معالم الفقرات ما بين المجموعات المختلفة (اي الفقرات التي تختلف بنماذج او خطوات الاجابة)، حيث ان هذه المشكلة الرئيسية في هذا النموذج هي مشكلة المقارنة ما بين تلك الفقرات وليس في تقدير الاجابات.

طرق معادلة الاختبارات باستخدام النظرية الحديثة في القياس التربوي:

هناك ثلاث طرق رئيسية لمعادلة الاختبار بهذه النظرية وهي:

(أ) معادلة الاختبارات باستخدام درجات القدرة (السِّمة) Ability Score Equating.

في هذا النوع منٍ معادلة الاختبارات نفترض أن كلا الاختبارين (X, Y) يقيسان نفس القُدرة (السمة) (θ)، لذلك فعند تقدير معلمة القدرة ومعالم الفقرات في نفس الوقت لكل من الاختبارين فإننا نضعهما على نفس التدريج؛ حيث $\theta_x = \theta_y$. وللقيام بهذه الخطوة فإننا نحتاج إلى تدريج درجات القدرة أثناء معايرة Calibration الاختبارات. وتعني المعايرة هنا تقدير معلمات الفقرات، كالصعوبة والتمييز. وينتج عن ذلك علاقة خطية بين تدريج القدرة وتقدير معلمات الفقرات. بشرط أن يكون التقديران منفصلين.(Kolen & Brennan, 1995).

(ب) معادلة الاختبارات باستخدام الدرجة الحقيقية True-Score Equating

في النظرية الكلاسيكية للاختبارات(CIT) ، يُطلَقُ على الدرجة المتوقعة للمفحوص (g) على الفقرة مصطلح "الدرجة الحقيقية". بينما في النظرية الحديثة في القياس التربوي والنفسي، تصاغ رياضياً على الصورة:

$$\sum_{1=1}^{n} P_1(\theta_g)$$

فلو افترضنا أن الاختبار بصورتيه يقيس نفس السمة، وأن معالم الفقرات لكلا الصورتين أو الاختبارين قد تم وضعهما على نفس التدريج؛ فإن الدرجة الحقيقية الصحيحة للطالب هي ζ (زيتا) في الاختبار X، وη (ايتا) في الاختبار Y؛ وتكونان مرتبطتين بدرجات القدرة (θ) بواسطة دالة خاصية الاختبار Test Characteristic Curve، بالصورة الرياضية الآتية .(Lord, 1980)

$$\eta = \sum_{1=1}^{n} P_1(\theta) \qquad \zeta = \sum_{1=1}^{n} P_1(\theta)$$

ويجب أن تكون تدريجات القياس للدرجات ζ η، مستقلة عن مجموعة المفحوصين n في الاختبار، وأن يكون تدريج قياس درجات القدرة مستقلاً عن عدد فقرات الاختبارين، كما يُشترط أن تكون مستوياتِ صعوبة الفقرات متطابقة في الاختبارين.

ومن عيوب معادلة الاختبار بهذه الطريقة، ان الدرجات الحقيقية المُقدّرة لا تناظر الدرجات الخام مناظرة واحد لواحد. والعيب الأخر هو أن الدرجات الحقيقية المُقدَّرة تُوضع على نفس التدريج مع الدرجات الخام (Peterson et. al., 1989).

(ج) معادلة الاختبارات باستخدام طريقة الدرجات المشاهَدَة (الدرجات الخام) Observed-Score Equating

من المشكلات التي يواجهها المختصون في معادلة الاختبار بطريقة الدرجات الحقيقية، أن هذه الطريقة لا تنتج عنها درجات معادلة للمفحوص الذي درجته الخام أدنى من مستوى الصدفة(التخمين) (C)؛ وذلك لان العلاقة بين الدرجات الخام ليست مثل العلاقة بين الدرجات الحقيقية. ففي الدرجات الخام تكون أدنى درجة هي الصفر، وفي الدرجات الحقيقية تكون أدنى درجة هي $\sum_{1=1}^{n}(C)_i$ ، أي درجة التخمين.

تقوم معادلة الاختبار باستخدام طريقة الدرجات الخام على فكرة التنبؤ بالتوزيع النظري للدرجات الخام للاختبار،عن طريق بناء التوزيع التكراري الذي تمثله الدالة $f(X \mid \theta)$ للدرجات الخام للاختبار لمفحوص قدرته (θ). فإذا وجدنا أن دوال الاستجابة لكل فقرة من فقرات الاختبار متطابقة، بحيث يكون $P_1(\theta) = P(\theta)$ ، فإن التكرار النسبي للدرجات الصحيحة (X) للمفحوص (g) يمكن حسابه رياضياً بالمعادلة الآتية ضمن توزيع ذي الحدين:

$$\int(X|\theta_g) = \binom{n}{x} p_x Q^{n-x}$$

كما يمكن معادلة الاختبار بهذه الطريقة للوصول إلى دالة التكرار النسبي ذي الحدين بإتباع الخطوات الآتية:

١. وضع معلمات القدرة ومعلمات الفقرات على تدريج مشترك لكل المجموعات والاختبارات.

٢. الحصول على التوزيع التكراري الهامشي Marginal Frequency Distribution للدرجات في الاختبار الأول باستخدام تقديرات المعلمات على الاختبار، وتقديرات معلمات القدرة، باستخدام الدالة الرياضية الآتية:

$$\int(X) = \sum_{1-1}^{n} \int(X|\theta_g)$$

٣- . تكرار الخطوة رقم (٢) للاختبار الثاني.

٤. إجراء معادلة للاختبار بطريقة الرتب المئينية المتساوية بين الدرجات الخام في الاختبار الأول ونظيراتها في الاختبار الثاني، وذلك باستخدام التوزيع التكراري الهامشي الذي تم إنشاؤه. ومن المهم جداً في معادلة الاختبار بهذه الطريقة تغطية مدى الدرجات الخام بالكامل.

ويوضح الجدول الآتي مقارنة بين طرق المعادلة المختلفة .(الدوسرس،٢٠٠٤)

جدول رقم (١)

طرق معادلة الاختبار وبعض خصائصها

طريقة المعادلة	الدالة	الاحصاءات المتشابهه	طريقة عرض النتائج
الوسط الحسابي	Y=X+B	الوسط الحسابي	كثابت تحويل (B) وقاعدة تقريب
الخطية	Y=AX+B	الوسط الحسابي والانحراف المعياري	كميل (A) وقاطع (B)، وقاعدة تقريب
الرتب المئينية المتساوية	معقدة	الوسط الحسابي والانحراف المعياري	تتطلب جدول تحويل للدرجات
IRT	معقدة	شكل توزيع الدرجات	تتطلب برناج حاسوبي لتحويل الدرجات

ثالثا: طريقة كيرنيل في معادلة الاختبارات(Kernel method) :

إن معادلة كيرنيل تشير إلى مجموعة جديدة من طرق المعادلة اللامعلمية والتي تستخدم التوزيع الطبيعي لمعادلة الدرجات من نموذج إلى آخر، فقد تم تطويرها من قبل باول هولاند ودورثي وتاير (Holland ,et al (1987) ولاحقا تم تحديثها من قبل ألينا فون ديفر.((Alina Von Daveir, et, al. 2004

تعتمد طريقة كيرنيل على مجموعة من طرق المعادلة الشبيهة بالمعادلة المئينية، وتنتمي إلى الطرق الكلاسيكية في معالجة الدرجات المشاهدة، ويمكن أن تستخدم طريقة كيرنيل مع مختلف تصاميم جمع المعلومات ومن ضمنها التصاميم التي تعتمد على وجود الجذع المشترك.(.(Alina Von Daveir,et,al. 2004

يوجد هناك طريقتان رياضيتان تستخدمان لمعادلة درجات الاختبارات باستخدام طريقة كيرنيل:

١- معادلة تشين Chain equating- CE -وفي معادلة تشين تواجه أولا الربط بين نموذج (x) مع الجذع المشترك A ، ثم يتبع بربط A مع نموذج (Y).

٢- معادلة ما بعد المطابقة (Post Stratification-PSE) حيث يتم تقدير التوزيع الهامشي لنموذج(x) و نموذج (Y) في المجتمع المستهدف ثم بعد ذلك تتم إجراءات المعادلة . (Alina Von Daveir,et,al. 2004)

لمعادلة الاختبارات ذات الفقرات المشتركة باستخدام طريقة كيرنيل، نحتاج الى خمس خطوات اساسية، والطريقة التي يتم فيها تنفيذ كل خطوة من هذه الخطوات الخمس تختلف باختلاف نوع الاختبار ذات الفقرات المشتركة الذي يستخدم هل هو: داخلي (Internal) أم خارجي(External) ، وتتلخص هذه الخطوات الخمس في النقاط الآتية:

الخطوة الأولى : قبل التمهيد Pre-smoothing

إن الهدف الرئيسي من هذه الخطوة هو ايجاد توزيع الدرجات الخام للقيام باجراءات المعادلة، كما أنها تهدف الى تكوين المصفوفة التي سوف تستخدم لاحقا لحساب الخطأ المعياري للمعادلة. حيث يتم في هذه الخطوة تقدير احتمالات الدرجة في تصميم المعادلة المستخدم، والتي تحدد مدى مطابقة النموذج الإحصائي المستخدم للعلامات الخام ، وهذا النموذج الإحصائي الذي يستخدم يختلف باختلاف التصميم المستخدم (Alina Von Daveir,et,al. 2004).

الخطوة الثانية: تقدير احتمالات الدرجة

(Estimation of score probabilities)

الخطوة الثالثة: الاستمرارية :Continuization

في هذه الخطوة يتم تحديد متوسط توزيع كيرنيل و متوسط (الدرجات المنفصلة)، ويترك التباين لتوزيع كيرنيل دون تحديد، لكن يتم التحكم بتباين كل توزيعات كيرنيل من خلال استخدام احد المعالم ويسمى (bandwidths, *hX* and *hY*).

الخطوة الرابعة: المعادلة Equating

بعد أن يتم توزيع الدرجات في نقاط محددة، تكون الخطوة التالية حساب معادلة كيرنيل وتتم هذه المعادلة بالصيغة الآتية لمعادلة (x) إلى (y):

$$e_y(x) = G^{-1}(F(x))$$

حيث ان

$G^{-1}(F(x)$: هي معكوس التوزيع التراكمي للنموذج (x)

ولمعادلة نموذج (y) إلى (x).(Alina Von Daveir,et,al. 2004)

$$e_y(y) = G^{-1}(F(Y))$$

حيث ان :

$G\text{-}1(F(Y))$: هي مقلوب التوزيع التراكمي للنموذج (Y)

الخطوة الخامسة: حساب الخطأ المعياري للمعادلة:

إن الخطأ المعياري للمعادلة يحدد مقدار الدقة في عملية المعادلة لصور الاختبار ، ويمكن حساب الخطأ المعياري لمعادلة $\bar{X}$ إلى $\breve{Y}$ كما يأتي:

$$SEE\, y_{(CE)}(X) = \sqrt{\left[SEE_y(e_A(x)\right] + \left[\bar{e}_y(e_A(x)SEE_A(x)\right]^2}$$

ويجب أن يلاحظ أن إيجاد الخطأ المعياري للمعادلة هو أكثر ملاءمة للعينات الكبيرة وكذلك يمكن أن يحسب في حالة العينات الصغيرة(Alina Von Daveir,et,al. 2004).

مشكلات فنية وتطبيقية في معادلة الاختبارات:

يقول أنجوف (Angoff (1980 أن أدبيات القياس والتقويم التربوي والنفسي قد اهتمت بتطوير نماذج لمعادلة الاختبار، ودوالها الخاصة بخطأ المعادلة، وكذلك مدى قوة وإحكام Robustness هذه النماذج في مواجهة عدم استيفاء متطلبات الفروض الأساسية لتطبيق وتطوير هذه النماذج.ومن المشكلات الفنية التي يواجهها المختصون عند القيام بمعادلة الاختبارات:

- معادلة الدرجات في الاختبارات غير المتكافئة في محتواها للحصول على درجات متكافئة:

يشير لندكويست (Lindquist (19678 إلى أن اجراء المعادلة لاختبارين تتطلب الحصول على درجات كلية متكافئة، بشرط الحصول على رتب مئينية متساوية للمجموعات التي طبقت عليها الاختبارات. ولكن في بعض الأحيان، نجد أن بعض الطلبة قد أخذوا أحد الاختبارين أو كليهما، مما يفسح المجال لتأثير عامل الزمن على درجات الطلبة. وخاصة إذا تزامن مع ذلك اختلاف الاختبارين المراد معادلتهما في المحتوى إلى حدٍ ما. مما يعرقل عملية إحلال اختبار محل آخر ضمن عملية تكافؤ الاختبارين. وفي مواقف أخرى، يرغب الباحثون أحياناً في استخدام نتائج تقويم اختبارات المواد في المرحلة الثانوية، بغرض التعرف على جوانب القوة والضعف في ابعاد شخصية الفرد Intra-Individual، وكذلك لمعرفة قيمتها الإرشادية، ثم جعلها متكافئة مع بعضها عن طريق معادلة فقراته (Lennon, 1967). والمشكلة في هذه الاختبارات أنه لا يمكن تطبيقها على مجموعة تقنين مفردة أو مشتركة. وجذور هذه المشكلة تعود إلى تقدم الطالب في التحصيل من صف إلى آخر. لذلك فإن أي رُتب مئينية يتم الحصول عليهما من نتائج هذه النوع من الاختبارات لا يمكن اعتبارها متكافئة، على الرغم من إمكانية التعامل مع هذه المشكلة من خلال وجود جذع مشترك Anchor Test. ويرتبط بهذه المشكلة عدة عوامل تجعل عملية معادلة الاختبارات أكثر صعوبة. وأبرز هذه العوامل هو ايجاد درجات متكافئة، وحجم العينة المسحوبة. وانخفاض ثبات درجات هذه الاختبارات، والطريقة المستخدمة في عملية المعادلة.

- أخطاء معادلة الاختبار: يهتم علماء القياس بمقارنة الطرق المستخدمة في عملية معادلة الاختبار، للتعرف الى اكثر هذه الطرق دقة في معادلة الاختبارات، وذلك للحصول على معادلة اختبارات عالية الثبات. فعلى سبيل المثال، أجرى جفني وميلامد (Gafni & Melamed, 1991) دراسة استخدما فيها الطرق الخطية في عملية المعادلة، لتقدير مقدار خطأ المعادلة لاختبار واحد. المقاييس المعيارية التي استخدمها الباحثان كانت متوسط الفرق بين الدرجة الحقيقية (الخام) والدرجة المعادلة، وكذلك مربع متوسط الجذر Root Mean Square لذلك الفرق. ولم يجد الباحثان علاقة ارتباطية واضحة بين خطأ معادلة الاختبار وبين طرق عملية المعادلة. أم برينان وكولن (1987) Brennan & Kolen فيشيران إلى أن القضايا العملية الحساسة في عملية معادلة الاختبار ترتبط بالتعرف على المصادر المتعددة لخطأ معادلة الاختبار؛ ثم حساب مقداره، والتخلص منه، أو خفضه إلى الحد الأدنى. ويرى الباحثان أن القضايا الحساسة في معادلة الاختبار هي توصيف المحتوى، ومعادلة الاختبار في سياق درجات القطع، وإعادة معادلة الاختبار، وأثر خرق عنصر الأمن والسرية في معادلة الاختبار. ويؤكدان بأن هذه القضايا لا يمكن معالجتها دون التواصل الدقيق والفعال بين المختصين في القياس والتقويم التربوي والنفسي، وخاصة فيما يتعلق بالحدود والقيود المرتبطة بأخطاء عملية المعادلة تلك.

- تأثير قدرة المفحوص في دقة معادلة الاختبار: كشفت البحوث التي أجريت على معادلة الاختبار باستخدام طرق نظرية استجابة الفقرة عن نتائج مختلفة فيما يتعلق بدرجة اللاتباين invariance. فقد قام سكاجز وليستز (Skaggs & Lissits (1988 بإجراء دراسة لمقارنة عدة طرق لمعادلة الاختبار(طريقة الرتب المئينية وطريقة راش، وطريقة IRT ثلاثية المعلم) لفحص لاتباين عملية المعادلة، وذلك عن طريق المعادلة الرأسية لاختبارين تحت ظروف مختلفة. وجد الباحثان أن تعددية البُعْدْ (السمة) Multidimensionality قد تكون السبب وراء اختفاء خاصية اللاتباين. واستخلص الباحثان أن مسألة تعددية السمة هي مسألة درجة degree. وفي الواقع من الصعب معرفة عند أية مسألة تعددية السمة هي مسألة درجة degree. وفي الواقع من الصعب معرفة عند أية نقطة يمكن استبعاد عملية المعادلة عندما لا يقيس الاختباران نفس السمة تماماً. ويرى برينان وكولن (Brennan & Kolen (1987 أن مسألة اللاتباين هي قضية مرتبطة بمجموعة جزئية من المفحوصين مستقاة من المجتمع الأصلي للدراسة.(الدوسري،٢٠٠٤)

• أثر طول الاختبارين واختلافهما في المحتوى على معادلة الاختبار: اهتم بهذه المشكلة كل من بيترسون وكوك وستوكنج Peterson, et al (1983). أنجوف Angoff (1980). فقد وجد بيترسون وزملاؤه أن استخدام نماذج IRT اللوجستية Logistic تؤدي إلى استقرار في نتائج معادلة الاختبارات التي تختلف في الطول (عدد الفقرات) وفي المحتوى، بشرط مراعاة بعض العوامل، مثل: ملاءمة النموذج المستخدم للبيانات المستخدمة في عملية المعادلة ، وعدد فقرات الربط Linking Items لوضع تقديرات معلمات الفقرات على نفس التدريج. وإذا استطاع الباحث الحصول على مجموعتين من المفحوصين متكافئتين عشوائياً، فإن استخدام اختبار مشترك بينهما يؤدي إلى نتائج دقيقة في معادلة الاختبار.

• فاعلية الطرق المستخدمة قبل معادلة الاختبار: يشير أنجوف Angoof (1980) إلى أن عملية معادلة الاختبار قبل تطبيقه (معادلة الصورة القديمة مع الصورة الجديدة للاختبار) تسمح بالحصول على مجموعة كبيرة من فقرات الاختبار المعايرة. وضمن السياق نفسه، أجرى كولن وهاريس Kolen & Harris (1990) دراسة لمعادلة الاختبار واستخدما تصميمين لهذا الغرض، هما تصميم المجموعات العشوائية وتصميم ما قبل المعادلة للفقرات؛ وذلك باستخدام طريقة الرتب المئينية، وطريقة IRT. وجد الباحثان أن عملية المعادلة قبل تطبيق الاختبار تعطي نتائج غير ذات فائدة عملية.

تقييم نتائج المعادلة

بعد أن يتم تطبيق المعادلة، فان نتائج المعادلة يمكن أن تقيم بمعايير متعددة، ومن هذه المعايير التي استخدمت في هذه الدراسة:

أولا: الخطأ المعياري للمعادلة Standard Error for Equating -(SEE)

من المشكلات الهامة التي يتصدى لها المختصون في معادلة الاختبارات مشكلة الخطأ الناجم عن عملية المعادلة نفسها، وكيفية تقليص هذا الخطأ إلى الحد الأدنى، وذلك لرفع مستوى دقة القياس إلى الحد الأقصى، وللحصول على معادلة اختبار صحيحة يمكن تفسير نتائجها تفسيراً علميا وعمليا بشكل صحيح.

يعكس الخطا المعياري في معادلة درجات الاختبارات الاختلاف بين درجات المعادلة المقدرة من العينة والدرجات المعادلة من المجتمع، حيث يتألف خطا المعادلة من مصدرين للخطأ هما:

١- خطا المعادلة العشوائي وهو خطا ناتج عن اختيار العينة. ويحدث عند تقدير بعض المعالم الاحصائية مثل (الوسط الحسابي، والانحراف المعياري وغيرهما) للعينات الصغيرة من مجتمع الدراسة، ويمكن تخفيض هذا النوع من الخطأ باستخدام عينات كبيرة من المفحوصين المطبقة عليهم معادلة الاختبار، ثم اختيار التصميم الأنسب لعملية المعادلة.

٢- أما الخطأ المنتظم لمعادلة درجات الاختبارات، فينجم عادة عن عدم التزام الباحث بالافتراضات والشروط التي تتطلبها طريقة المعادلة، فمثلاً في تصميم المجموعات المفردة، إذا أخفقنا في ضبط الآثار الناجمة عن عاملي التعب، والخبرة والممارسة،وهما مصدران أساسيان للخطأ المنتظم، نكون قد قللنا من دقة وصحة معادلة الاختبار. كذلك في تصميم المجموعات العشوائية، ينتج الخطأ المنتظم نتيجة لعدم تفاوت قدرات مجموعة المفحوصين، أما في تصميم المجموعات ذات الفقرات المشتركة، فنجد ان الخطأ المنتظم يأخذ في الزيادة إذا كانت الفقرات المشتركة تلك لا تمثل الاختبار ككل في محتواه وفي خصائصه الإحصائية. كذلك تزيد قيمة الخطأ المنتظم إذا كان تأثير الفقرات المشتركة مختلفاً في صورتي الاختبار. ويمكننا القول بشكل عام، أن أي تصميم من تصميمات معادلة الاختبار يُنْتِجُ الخطأ المنتظم إذا كانت صورتا الاختبار مختلفتين في المحتوى، وفي مستوى صعوبة الفقرات وفي الثبات (الدوسري، ٢٠٠٤).

ويتم حساب الخطا المعياري لدرجات الاختبار وعند كل فقرة كما يلي:

$$SEE = \sqrt{\left[\frac{1}{n-1}\sum_{1}^{n}(\hat{e}_x(y_k) - \bar{e}_x(y_k)\right]^2}$$

حيث:

n = عدد مرات التكرار.

y_k = تمثل الدرجات على النموذج y.

$e_x(y_k)$ = هي الدرجة المعادلة على النموذج x من خلال النموذج Y .

$\bar{e}_x(y_k)$ = هي متوسط درجات المعادلة للدرجة y_k عبر مرات التكرارات.

(Kolen,Brenann,2004

الجذر التربيعي لمتوسط مربعات الأخطاء ((RMSE)) :

كما ذكرنا سابقا فان الأخطاء المعيارية للمعادلة تعكس الأخطاء العشوائية للمعادلة، في حين يعتبر (RMSE) مهما لأنه يعكس مقدار التحيز للفقرة، وكذلك يعكس دقة نتائج المعادلة مقارنة بمعيار المعادلة المستخدم.

ويمكن حساب جذر متوسط مربع الأخطاء المعيارية (RMSE) من خلال المعادلة الآتية :

$$RMSE = \left[\frac{1}{n_s} \sum \frac{1}{n_e} \sum_j (y_j - E_j)^2 \right]^{\frac{1}{2}}$$

حيث أن :

n = عدد المفحوصين.

j = رمز المفحوص.

n_s = عدد العينات والمتمثلة في عدد مرات توليد البيانات .

n_e = عدد المفحوصين في العينة الواحدة.

y_j = الدرجة على الصورة الأولى للمفحوص j في ضوء درجته على الصورة الثانية.

E_j = الدرجة الحقيقية المتوقعة للمفحوص على الصورة الأولى.

ثانيا:الدراسات السابقة:

اجريت عدة دراسات لمقارنة طريقة كيرنيل في معادلة درجات الاختبارات،منها:

دراسة وانج (Wang, 2005) والتي هدفت إلى مقارنة الطريقة اللوغارتميه الخطية المستمرة مع طريقتين في المعادلة: طريقة كيرنيل وطرق المعادلة المئينيه غير الممهدة، تم استخدام تصميم المجموعات العشوائية لجمع البيانات من اختبار تم تطبيقه على الطلبة، وذلك لمقارنة طرق المعادلة الثلاث.أشارت النتائج أن طريقة المعادلة اللوغارتمية الخطية المستمرة أنتجت توزيعا مستمرا أكثر تمهيدا، كذلك عندما تمت مقارنة الطريقتين:معادلة كيرنيل اللوغارتميه (المستمرة) مع طريقة المعادلة المئينيه غير الممهدة وباستخدام نفس البيانات، أشارت النتائج أن الطرق الثلاث أظهرت نفس نتائج المعادلة.

ودراسة ماو (Mao, 2006) والتي هدفت إلى فحص دقة تقديرات الأخطاء المعيارية لمعادلة كيرنيل تحت ظروف مختلفة من: حجم العينة ودرجة التمهيد واختيار (bandwidth) وخصائص توزيعات الدرجة، تم جمع المعلومات باستخدام تصميم المجموعات العشوائية من خلال توليد بيانات لتقدير الخطأ المعياري للمعادلة ((SEE)). أشارت النتائج أن دقة تقدير الخطأ المعياري في المعادلة ((SEE)) كان أفضل في العينات ذات الحجم الكبير، كذلك أشارت النتائج أن الدرجات العالية من التمهيد تميل إلى إنتاج حجم اكبر من الأخطاء المعيارية للمعادلة مقارنة مع الدرجات المنخفضة من التمهيد.

واشارت دراسة ليفينجستون (Livingston, 1993): والتي هدفت إلى فحص دقة معادلة كيرنيل مع أحجام عينات صغيره باستخدام تصميم المجموعات غير المتكافئة ذات الاختبار المشترك، تم استخدام استجابات أكثر من (٩٣٠٠٠) مفحوص على اختبار مكون من (١٠٠) فقرة. تم تطبيق المعادلة المئينيه غير الممهدة على توزيعات المجتمع للحصول على معايير المعادلة، بعد ذلك تم سحب عينات عشوائية من المفحوصين لتطبيق معادلة كيرنيل عليها. وقد تم تكرار هذه العملية(٥) مرات بأحجام عينات مختلفة (٢٥-٥٠-١٠٠-٢٠٠). أشارت النتائج أن معادلة كيرنيل كانت أكثر دقه في معادلة الدرجات مقارنة مع النتائج التي نتجت عن معادلة تشين المئينيه للتوزيعات الممهدة وغير الممهدة.

واشارت دراسة كو(Qu, 2007) والتي هدفت إلى معرفة اثر البيانات المركبة من مجموعتين تحت تأثير اختلاف حجم العينات على دقة المعادلة. استخدم الباحث طريقتين في معادلة الدرجات وهي: طريقة كيرنيل في المعادلة التي تعتمد على المعادلة المئينيه والطريقة التقليدية في المعادلة. طبق الباحث نوعين من البيانات: بيانات حقيقية، وبيانات مولده ، ثم قام الباحث بإيجاد كل من الخطأ المعياري للمعادلة، البواقي المعيارية، تحيز المعادلة، والخطأ المعياري لاختلاف المعادلة لكل طريقة من طرق المعادلة المستخدمة.أشارت النتائج أن طريقة المعادلة باستخدام البيانات المركبة من مجموعتين تكون أكثر حساسية للظروف المختلفة من البيانات غير المركبة.

واشارت دراسة أكور (Akour, 2006): والتي هدفت إلى المقارنة بين ثلاث طرق لمعادلة الاختبارات باستخدام طرق المعادلة المئينيه وهي: طريقة ما قبل التمهيد، طريقة المعادلة الخطية المتعددة الحدود لما قبل التمهيد، وطريقة كوبيك سبلين بعد التمهيد، وبين طريقتين مختلفتين من طرق كيرنيل في المعادلة: (طريقة) كيرنيل، وطريقة المعادلة الخطية المستمرة (CLL). قام الباحث بجمع البيانات باستخدام تصميم المجموعات المتكافئة لتقدير معالم الفقرات باستخدام برمجية حاسوبية للحصول على عينات عشوائية من مجتمع التوزيعات، ومن ثم تطبيق إجراءات المعادلة على هذه العينات، حيث اعتمدت الدراسة على ثلاثة متغيرات رئيسية وهي: حجم العينة ومستوى الصعوبة وطول الاختبار. طبق الباحث تقنيات نظرية (IRT) لتمهيد التوزيعات الملاحظة؛ وذلك ليلائم توزيعات المجتمع المستهدف، بعد ذلك تم تطبيق المعادلة المئينية على التوزيعات الممهدة للحصول على علاقات المعادلة في المجتمع. أشارت نتائج الدراسة أن جميع طرق المعادلة كانت أكثر فاعلية من طرق قبل التمهيد.

اما بالنسبة لدراسات معادلة الاختبارات باستخدام نظرية (IRT)، فكانت عديدة، نذكر منها:

دراسة الصمادي (٢٠٠٦): والتي هدفت إلى الكشف عن فاعلية طرق تصحيح الصواب والخطأ المتعدد وتأثيرها على دقة معادلة الاختبارات باستخدام نماذج النظرية الحديثة المتعددة الاستجابة، ولتحقيق هذا الهدف قام الباحث ببناء اختبار تحصيلي مؤلف من (٣٥) فقرة لكل منها أربعة بدائل من نوع فقرات الصواب والخطأ المتعدد في مبحث الرياضيات.

تكونت عينة الدراسة من (٨٧٣) طالبا وطالبة من الصف الأول الثانوي موزعين على عشرة مدارس تشمل تسعا وعشرين شعبة في تخصصات الأدبي والعلمي والإدارة المعلوماتية للعام الدراسي(٢٠٠٤ / ٢٠٠٥) . أظهرت النتائج أن طريقة التصحيح الرابعة –وهي إعطاء الطالب علامة واحدة لكل بديل تم الإجابة عليه بشكل صحيح – والتي تعتمد على مراعاة المعرفة الجزئية، كانت الأكثر دقة في قياس قدرات الأفراد، وتعطي معلومات أكثر عن الاختبار، كما كانت الأكثر فاعلية في معادلة الاختبار.

وهدفت دراسة يوكي(Yuki, 2008) إلى مقارنة الطرق المعلمية واللامعلمية في نظرية استجابة الفقرة في معادلة الدرجات باستخدام تصميم المجموعات غير المتكافئة ذات الجذع المشترك. تم استخدام أربع طرق معلميه لمعادلة الدرجات في نظرية استجابة الفقرة: الدرجات الحقيقية والملاحظة المعتمدة على تقدير معالم الفقرات، بالإضافة إلى استخدام أربع طرق لامعلمية وهي: معادلة الدرجات الحقيقية والمشاهدة بالاعتماد على توزيع كيرنيل. وحتى تتم المقارنة في الأداء بين الطرق المستخدمة تم الاعتماد على نوعين من البيانات: حيث تم توليد بيانات بالاعتماد على خمسة عوامل :عدد الفقرات، نسبة الفقرات التي تنتهك افتراضات النموذج المعلمي، عدد المفحوصين، توزيعات القدرة لمجموعات المفحوصين، وأخيرا فيما إذا كانت العلامات لها معلم ذات خط تقاربي منخفض أم لا، كذلك تم تطبيق بيانات حقيقية لفحص سلوك طرق المعادلة في المواقف الحقيقية. أشارت نتائج البيانات المولدة والبيانات الحقيقية انه عندما تتحقق افتراضات النموذج، تكون الطرق المعلمية أكثر دقة من الطرق اللامعلمية، كذلك الطرق اللامعلمية تصبح أكثر دقة من الطرق المعلمية عندما تنتهك افتراضات النموذج، ومن النتائج الأخرى التي توصلت إليها الدراسة أن الخط التقاربي الأسفل يؤثر في نتائج المعادلة، بالإضافة إلى أن الطرق اللامعلمية هي أكثر دقة مع الدرجات المشاهدة من الدرجات الحقيقية في معادلة الدرجات.

واشارت دراسة روبرت (Robert,2007)التي هدفت إلى مقارنة الدرجات الحقيقية في نظرية استجابة الفقرة ونظرية استجابة الفقرة بالاعتماد على المعادلة الموضعية، قام الباحث باستخدام اختبارات مختلفة وبنماذج مختلفة لفحص الأداء وذلك لتقييم نتائج معادلة الدرجات الحقيقية مع معادلة الدرجات المشاهدة في نظرية استجابة الفقرة تحت ظروف مختلفة: طول الجذع المشترك، فقدان بعض البيانات، طريقة التدريج، وتوزيع قدرات المفحوصين. تم تقييم نتائج المعادلة بالاعتماد على البواقي المعيارية في المعادلة وا لتحيز. أشارت

النتائج أن تقديرات الدرجات الحقيقية أظهرت تقديرات اقل في معياري التحيز والبواقي المعيارية، كذلك لم توفق الدراسة للإشارة بوضوح لاختيار طريقة التدريج لكلا طريقتي المعادلة.

واشارت دراسة يونغ وو (Young Woo,2007) التي هدفت إلى مقارنة الخطأ المعياري للمعادلة باستخدام طرق نظرية استجابة الفقرة والمئينات مع فقرات متعددة الاستجابة باستخدام تصميم المجموعات غير المتكافئة ذات الاختبار المشترك. استخدمت هذه الدراسة بيانات حقيقية من اختبار مخصص لتقييم الكتابة، حيث عدلت البيانات الأصلية لعمل خمس صور من الاختبارات وثلاث صور أخرى للاختبار، وتم تطبيق معايير الخطأ المعياري والبواقي المعيارية على طرق المعادلة الستة المستخدمة في هذه الدراسة، وبحجم عينات(ن=٢٥٠-٥٠٠-١٠٠٠-١٦٨٠) للاختبارات الخمسة وتم حسابه باستخدامBootstrap، وبأحجام عينات(٥٠٠-١٠٠٠- ١٦٨٠) للاختبارات الثلاثة.
أشارت النتائج بصورة عامة أن طريقة المعايرة المتلاقية أظهرت أخطاء معيارية وبواقي معيارية اقل من طريقة المعايرة المنفصلة، كذلك أظهرت النتائج أن معادلة الدرجات المشاهدة أنتجت أخطاء معادلة وبواقي معيارية اقل من معادلة الدرجات الحقيقية.

اما دراسة هيوجين (Huigin,2004) والتي هدفت إلى التحقق من طرق المعادلة في نظرية استجابة الفقرة مع وجود القيم الشاذة. تم توليد بيانات تجريبية باستخدام تصميم المجموعات غير المتكافئة ذات الجذع المشترك وتحت ظرف اختلاف قدرة المفحوصين(المجموعات المتكافئة ن(٠،١) مقابل ن (١،١). كما تم تطبيق عشر طرق من طرق المعادلة في نظرية استجابة الفقرة لمعادلة البيانات التي تم توليدها تجريبيا. ولتقييم دقة هذه الطرق في المعادلة تم استخدام الأخطاء المنتظمة. أشارت النتائج انه عندما تكون القيم الشاذة متضمنة في البيانات، فان أداء الطرق العشرة المستخدمة في المعادلة تؤدي إلى ظهور اختلافات في قدرة المفحوصين وفي توزيع نقاط الدرجة للقيم الشاذة.

اما بالنسبة للدراسات التي اجريت لمقارنة طريقة المئينات في معادلة الاختبارات فنذكر منها:

دراسة هاريس و كولن (Harris, Kolen, 1990): والتي هدفت إلى معرفة مدى فاعلية طرق المعادلة باختلاف مستويات القدرة، وكانت طرق المعادلة المستخدمة هي: الطريقة

الخطية، والطريقه المئينية الكلاسيكية، والمعادلة وفق النموذج الثلاثي المعلمة في نظرية استجابة الفقرة، واستخدمت لتحقيق هذا الهدف خمس صور لاختبارات المفاهيم الحسابية بحيث كانت متماثلة في درجة ثباتها، ودرجة صعوبتها والمحتوى. أشارت النتائج إلى عدم تميز أي من طرق المعادلة المستخدمة في إعطاء نتائج أفضل.

واشارت دراسة أيوب (1994) والتي هدفت إلى المقارنة بين أربع طرق لمعادلة الاختبارات وهي: الطريقة الخطية والطريقة المئينية المنبثقتان عن النظرية التقليدية في القياس ونماذج النظرية الحديثة في القياس أحادي المعلمة وثنائي المعلمة عندما يكون التصميم (أ) مجموعات متطابقة أو عشوائية، وعندما يكون التصميم(ب) مجموعات غير متكافئة مع فقرات مشتركة. لتحقيق هدف الدراسة تم بناء ثلاثة اختبارات بصورتين (أ) و (ب) لمادة الرياضيات للصفوف الرابع والخامس والسادس .أشارت نتائج المعادلة الأفقية أن نماذج النظرية الحديثة في القياس كانت أكثر فاعلية من طريقتي المعادلة الخطية.كذلك أشارت نتائج المعادلة العمودية إلى أن الطريقة المئينية كانت أكثر فاعلية من الطرق الأخرى المستخدمة في الدراسة يليها النموذج الثنائي المعلمة ثم النموذج أحادي المعلمة.

كما اشارت دراسة لي وكولن وفرزبي (Lee, Kolen, Frezby, 2001) والتي هدفت إلى المقارنة بين معادلة اختبارين أحداهما ثنائي التدريج والآخر متعدد التدريج وفق نماذج النظرية الحديثة والنظرية الكلاسيكية في المعادلة، حيث تم استخدام طرق المعادلة التقليدية الآتية لمعادلة الاختبار الثنائي التدريج، وهي (الأوساط الخطية والمئينية)، بينما تم استخدام طرق المعادلة وفق النظرية الحديثة وهي طريقتي المعادلة التاليتين((GRM) و (NM)) في معادلة الاختبار المتعدد التدريج، وتم استخدام محك الجذر ألتربيعي للأوساط غير الموزونة للحكم على فاعلية الطريقة المستخدمة في المعادلة. أظهرت النتائج تشابهاً بين طرق معادلة العلامات الحقيقية والملاحظة في الاختبار المتعدد التدريج وفق نماذج النظرية الحديثة، والطرق الكلاسيكية المستخدمة في المعادلة (الأوساط الخطية والمئينية) في الاختبار ثنائي التدريج للعلامات الحقيقية.

واشارت دراسة هاو(Hue, 2004) والتي هدفت إلى معرفة أثر اختلاف طرق اختيار الفقرات في اختبار الجذع المشترك على دقة معادلة الاختبار التكيفي. حيث أجريت الدراسة

على طلبة جامعة أوهايو الأمريكية من ذوي الثقافات المختلفة كالصينية والكورية والاسبانية وهدفت إلى التعرف إلى أثر اختلاف طرق اختيار فقرات الجذع المشترك على دقة عملية معادلة الدرجات وذلك من خلال مقارنة القسم اللفظي للاختبار بالقسم غير اللفظي واستخدم الباحث طريقتين لاختيار فقرات الجذع المشترك هما طريقة معالم الفقرة (الصعوبة والتميز)وطريقة دالة الفقرة التفاضلية.وهدفت أيضاً إلى التعرف فيما إذا كانت إحدى الطريقتين أفضل من غيرها في تحديد دقة المعادلة.دلت النتائج أن اختيار فقرات الجذع المشترك باستخدام طريقة معالم الفقرات كانت أفضل من حيث الدقة.

واشارت دراسة الصمادي (٢٠٠٧) والتي هدفت إلى التعرف إلى دقة المعادلة المئينية العمودية من خلال استخدام تصميم الجذع المشترك للمجموعات غير المتكافئة، وذلك من خلال الخطأ المعياري للمعادلة المئينية العمودية . وقد استخدم الباحث طريقة المعادلة العمودية المنبثقة عن النظرية التقليدية في القياس، وخلصت الدراسة إلى أن دقة المعادلة المئينية العمودية كانت أعلى ما يمكن عندما كان عدد فقرات اختبار الجذع المشترك أكثر من ٢٠ فقرة وممثلة للمحتوى والصعوبة والتمييز، وأن توفر خاصية التمييز في الفقرات لاختبار الجذع المشترك تعطي دقة أكبر مما لو توفرت خاصية الصعوبة في الفقرات.

واشارت دراسة فيكتور (Victor, 2007) والتي هدفت إلى فحص دقة المعادلة باستخدام عينات صغيرة الحجم وباستخدام تصميم المجموعات العشوائية عندما تختلف الاختبارات في متوسط الصعوبة وبمستويات مختلفة من الصعوبة ، كذلك كان الهدف الثاني من الدراسة هو التحقق من اثر خطا المعادلة في تصنيف الأخطاء عندما تقدم النتائج تبعا للأداء المعياري. تم تحديد دقة المعادلة من خلال فحص مقدار الإحصائيات الثلاث الآتية :البواقي المعيارية و الخطأ العشوائي والتحيز. تم تطبيق تسع طرق للمعادلة على بيانات مولدة، وبمستويات مختلفة من الصعوبة، وبستة أحجام من العينات (٢٥-٥٠-٧٥-١٠٠-١٥٠-٢٠٠). أشارت النتائج المتعلقة بطول الاختبار وتصميم المعادلة المستخدم في هذه الدراسة أن دقة المعادلة في العينة صغيرة الحجم يعتمد على: اختلافات الصعوبة بين نماذج الاختبار، مدى الدرجات التي تقيم طريقة المعادلة وفقا له، حجم العينة، وطريقة المعادلة التي تطبق، كذلك أشارت النتائج أن الاختبارات المتوسطة وذات الصعوبة الكبيرة ومن خلال التركيز على خطا العينة لتحديد فيما إذا كان سيتم تطبيق المعادلة على أحجام عينات

صغيرة هو أمر غير مفيد ، في حين أن العينات الصغيرة أنتجت أخطاء في العينات بشكل كبير.

كذلك اشارت دراسة اماندا (Amanda,2008) هدفت هذه الدراسة إلى مقارنة طرق المعادلة في النظرية الكلاسيكية ونظرية استجابة الفقرة باستخدام تقييم عدد الدرجات الصحيحة أو إحراز الصيغة. حيث تم تطبيق سبع طرق مختلفة لمعادلة هذين النوعين من الدرجات وهي: ثلاث طرق كلاسيكية (طريقة توكر الخطية ،الطريقة غير الممهدة،وطريقة تشين المئينية)، وأربع طرق من طرق معادلة الدرجات في نظرية استجابة الفقرة (النموذج الأحادي ،النموذج الثنائي، النموذج الثلاثي، ونماذج الاختيار من متعدد).تم مقارنة الطرق السبعة المستخدمة في هذه الدراسة باستخدام بيانات حقيقية والتي تم جمعها تحت تعليمات SAT وبيانات مولدة، فالبيانات الحقيقية استخدمت فيها بيانات إحراز الصيغة، بعد ذلك تم استخدام نماذج الاختيار من متعدد حتى يكمل الاستجابات المحذوفة للبيانات المولدة ليتبع تعليمات إحراز الدرجات الصحيحة. في البيانات المولدة تم تطبيق نموذج الاختيار من متعدد لتوليد البيانات بحيث تمثل كلا التطبيقين. أشارت النتائج أن الطريقة التي أنتجت اقل قيمة للتحيز في البيانات الحقيقية والمولدة هي طريقة توكر الخطية.

من مجمل الدراسات السابقة يمكن القول أن هذه الدراسات تناولت معادلة الدرجات من جوانب متعددة، فمنها من تناول معادلة الدرجات باستخدام الاختبارات المشتركة الداخلية أو الخارجية مع التصاميم المختلفة لجمع المعلومات، وبعضها حاول التحقق من دقة معادلة طريقة من بين طرق المعادلة المختلفة.

وتطرقت بعض الدراسات إلى الحديث عن العوامل التي تؤثر في دقة معادلة الدرجات من مثل تلك المتعلقة بصعوبة الفقرات ،أو حجم العينات ،أو طبيعة البيانات المستخدمة، كما عالجت بعض الدراسات اثر زيادة كل من عدد فقرات الاختبار وحجم العينة على دقة معادلة الدرجات. وقد لاحظ الباحث أن بعض هذه الدراسات عانت من بعض جوانب القصور على الرغم من أن بعض هذه الجوانب تم الإشارة إليها كمحددات في هذه الدراسة مثل اقتصار عينتها على فئة معينة من الفئات المختلفة، إضافة إلى اقتصار بعض الدراسات على طريقة واحدة أو اثنتين من طرق معادلة درجات الاختبار المختلفة. وقد لاحظ الباحث أيضا أن الدراسات العربية رغم محدودية عددها لم تشمل طريقة كيرنيل

،حيث ركزت معظمها على معادلة الدرجات باستخدام المئينات أو طرق نظرية استجابة الفقرة.

ولعل هذه الدراسة تتميز عن الدراسات السابقة في كونها تناولت الدرجات متعددة الحدود في معادلة الدرجات. كذلك تعتبر هذه الدراسة استمرارا للدراسات التي تناولت متغيرات عديدة مثل حجم العينة، صعوبة الفقرات، وعدد فقرات الاختبار للوقوف على أفضل طرق معادلة الدرجات دقة وثباتا.

كما ان هذه الدراسة اعتمدت على نظرية استجابة الفقرة كمعيار للمقايسة والمفاضلة بين طرق المعادلة المختلفة، وذلك أن معظم الدراسات السابقة كانت تستخدم الطريقة المئينية للمفاضلة بين طرق المعادلة المختلفة.

الفصل الثالث

الطريقة والإجراءات

الفصل الثالث

الطريقة والإجراءات

هدفت هذه الدراسة إلى تقييم دقة معادلة صور الاختبار باستخدام طرق :كيرنيل، المئينات، وIRT في ظروف متنوعة من حالات طول الاختبار ، وحجم العينة ، ومستويات الصعوبة ، وفيما يلي وصفا للمنهجية المتبعة في هذه الدراسة.

أولا : تصميم الدراسة ومتغيراتها :

تكونت متغيرات الدراسة المستقلة من :

١- طرق المعادلة.

٢- طول الاختبار

٣- حجم العينة

٤- مستويات الصعوبة .

طول الاختبار : ويقصد به عدد الفقرات في الاختبار الواحد بما فيها فقرات الجذع المشترك ، وتكون من ثلاث مستويات :

١- ٣٠ فقرة .

٢- ٤٠ فقرة .

٣- ٦٠ فقرة .

حجم العينة: ويقصد به عدد الأفراد في العينة الواحدة ، وتكونت من ثلاثة مستويات :

١- ٢٠٠مفحوص.

٢- ٦٠٠ مفحوص.

٣- ١٠٠٠ مفحوص.

مستويات الصعوبة ، وله مستويان :

١- التشابه في متوسط معدل الصعوبة للاختبار.

٢- الاختلاف في متوسط معدل الصعوبة (٠,١) انحراف معياري.

وفي ضوء المتغيرات الثلاثة بمستوياتها ،كان تصميم الدراسة (٣ X ٣ X ٢) وبعدد خلايا (١٨) خلية ، والجدول الآتي يبين التصميم وخلاياه:

جدول رقم (٢)

خلايا التصميم متضمنا المتغيرات المستقلة في الدراسة وعدد مرات التكرار في كل خلية

طول الاختبار	طريقة المعادلة	التشابه في الصعوبة			الاختلاف في الصعوبة		
		٢٠٠	٦٠٠	١٠٠٠	٢٠٠	٦٠٠	١٠٠٠
30	المئينات	√	√	√	√	√	√
	IRT	√	√	√	√	√	√
	كيرنيل	√	√	√	√	√	√
40	المئينات	√	√	√	√	√	√
	IRT	√	√	√	√	√	√
	كيرنيل	√	√	√	√	√	√
60	المئينات	√	√	√	√	√	√
	IRT	√	√	√	√	√	√
	كيرنيل	√	√	√	√	√	√

ثانيا: المتغيرات التابعة :

وهو معيار الحكم على دقة معادلة درجات الصور المختلفة للاختبار وذلك بحساب الخطأ المعياري للمعادلة(SEE) ، والجذر التربيعي لمتوسط مربعات الخطأ (RMSE).

ويحسب الخطأ المعياري للمعادلة بالمعادلة الآتية :

$$SEE = \sqrt{\left[\frac{1}{n-1}\sum_{1}^{n}(\hat{e}_x(y_k) - \bar{\hat{e}}_x(y_k)\right]^2}$$

حيث:

n = عدد مرات التكرار.

y_k = تمثل الدرجات على النموذج y.

$e_x(y_k)$ = هي الدرجة المعادلة على النموذج x من خلال النموذج y_k .

$\bar{\hat{e}}_x(y_k)$ = هي متوسط درجات المعادلة للدرجة y_k عبر مرات التكرارات.(الحربي،٢٠٠٩)

الجذر التربيعي لمتوسط مربعات الخطأ

Root Mean Squared Error (RMSE)

ويمكن حساب الجذر التربيعي لمتوسط مربعات الخطأ(RMSE) من خلال المعادلة الآتية :

$$RMSE = \left[\frac{1}{n_s}\sum\frac{1}{n_e}\sum_{j}(y_j - E_j)^2\right]^{\frac{1}{2}}$$

حيث أن :

n= عدد المفحوصين.

j = رمز المفحوص.

n_s = عدد العينات والمتمثلة في عدد مرات توليد البيانات .

n_e = عدد المفحوصين في العينة الواحدة.

y_j = الدرجة على الصورة الأولى للمفحوص ز في ضوء درجته على الصورة الثانية.

E_j = الدرجة الحقيقية المتوقعة للمفحوص على الصورة الأولى.

وحسب مربع انحراف الدرجة المعدلة للمفحوص على الصورة الأولى عن درجته الحقيقية المتوقعة ، ومن ثم تم إيجاد مجموع مربعات انحرافات كل أفراد العينة في كل مرة ، وحسب متوسط مجموع مربعات الانحرافات في كل عينة ، ومن ثم متوسط عدد مرات التوليد وهي (٢٠٠) مرة، وبعد ذلك حسب الجذر التربيعي لمتوسط مربعات الخطأ.

توليـــــــد بيــــــانات الدراســـــة

تم توليد البيانات ومعادلة الدرجات وتقييم دقة المعادلة وفق الخطوات الآتية:

١- النموذج الذي تم اعتماده لتوليد البيانات .

تم توليد بيانات الدراسة باستخدام نموذج التقدير الجزئي المعمم الذي يتم من خلاله حساب احتمالية الاستجابة الصحيحة على الفقرة عند مستوى القدرة (θ) وفقا للمعادلة التالية :

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^{x} \alpha_i(\theta - \delta_{ij})\right]}{\sum_{j=0}^{M}\left[\exp\left[\sum_{j=0}^{r} \alpha_i(\theta - \delta_{ij})\right]\right]}$$

حيث أن:

(δ_{ij}) تمثل النقاط على متصل القدرة والتي تكون الإجابة فيها على فئة معينة اعلى من سابقتها في هذا النموذج(معامل صعوبة الفئة).

:ai: تمييز الفقرة.

Θ: هي قدرة المفحوص

٢- توليد قدرات الأفراد:

ولدت (Θ) والممثلة لقدرات أفراد كل عينة باستخدام برمجية (WINGEN2)

حيث ولدت (Θ) بتوزيع سوي متوسطه صفر وانحرافه المعياري واحد .

٣- خصائص الفقرات

تم تحديد خصائص الفقرات لجميع فقرات الاختبارات، وقد تم اختيار قيم معالم الفقرات كالتالي :

-معلم الصعوبة (b): تم اختيار قيم معالم الصعوبة للفقرات بحيث يكون لها توزيع سوي متوسطه صفر وانحرافه المعياري واحد N(0,.1) .

- معلم التمييز (a):تم اختـــــار قيم معالم التمييــــز للفقــرات بحيث يكون لها توزيع منتظم تتراوح قيمه بين (٠,٦٤ -١,٥) U(0.64 – 1.5) .

٤- **إجراءات توليد البيانات**

- تم توليد قيم θ لكل من الاختبارين وفقا لكل من توزيع القدرة وحجم العينة .
- تم توليد الاستجابات على الفقرات لكل قيم θ وفقا لمعالم الفقرات (الصعوبة والتمييز) وكذلك تم توليد استجابات للفقرات وفقا لنموذج التقدير الجزئي المعمم وبما يتلاءم مع معالم الفقرات.
- تمت إعادة الخطوة الأولى والثانية والثالثة ٢٠٠ مرة لكل حالة من الحالات

٥- **تكرار الخطوة السابقة بتوليد الاستجابات للمفحوصين باستخدام معالم الفقرات ومعالم قدرات المفحوصين ٢٠٠ مرة في كل خلية من خلايا التصميم .**

كررت الخطوة السابقة، وذلك باستخدام نفس البرمجية، حيث كان يتم في كل مرة توليد قدرات الأفراد على التوزيع المطلوب ، وباستخدام نفس معالم الفقرات لتحديد استجابات المفحوصين وفق النموذج المتعدد الاستجابات.

٦- **تخزين استجابات المفحوص في كل عينة وفي كل مرة تكرر فيها للصورة الأولى والثانية في الاختبار**

٧- **تم تحويل الملفات إلى ملفات نصية لصورتي الاختبار الأولى والثانية ؛ وذلك من اجل التعامل مع البيانات ، وتم إجراء المعادلة وفق الخطوات الآتية:**

* تم حساب درجة كل مفحوص على الصورة الأولى من الاختبار ككل ، ودرجته على فقرات الجذع المشترك.

* تم حساب درجة كل مفحوص على الصورة الثانية من الاختبار ككل ، ودرجته على فقرات الجذع المشترك.

* تم حساب المتوسط الحسابي والانحراف المعياري لأفراد العينة ،وذلك للدرجة الكلية وللدرجة على فقرات الجذع المشترك في الصورة الأولى ، وكذلك الأمر في الصورة الثانية ، ثم تم إجراء المعادلة وفق الخطوات الآتية:

أ) المعادلة المئينية المبنية على تصميم الاختبار المشترك، وتم إجرائها وفقا للخطوات الآتية:

١-إعداد توزيع العلامات لاختبار الجذع المشترك (u) للمجموعة الكلية والتي تشمل مجموعة الأفراد الذين أجابوا على الصورتين X, Y.

٢-إيجاد العلامات على الاختبار (u) في توزيع المجموعة الكلية التي تناظر مئينات معينة مثل (p5,p10,p15…p95).

٣-إعداد التوزيعات المنفصلة للاختبار المشترك (u) لكل مجموعة جزئية أجابت عن إحدى صورتي الاختبار المراد معادلتهما.

٤-يتم تحديد الرتب المئينية المناظرة للعلامات التي تم تعريفها في الخطوة رقم (٣) وذلك لكل مجموعة جزئية.

٥-إيجاد العلامة X أو Y المناظرة للرتبة المئينية التي حسبت في الخطوة (٤).

٦-تعد العلامتان X, Y متكافئتين حين تناظر كل منهما الرتبة المئينية ذاتها على اختبار الجذع المشترك والتي حددت من المجموعة الكلية التي أعطيت الاختبارين.

٧-رسم منحنى ممهد يمثل النقاط التي عينت في الخطوة رقم (٦) والذي يستفاد منه في تحديد التكافؤ بين صورتي الاختبار Y,X .

المعادلة في النظرية الحديثة للقياس:

تمت طريقة المعادلة في النظرية الحديثة في القياس بإتباع الخطوات الآتية:

١. وضع معلمات القدرة ومعلمات الفقرات على تدريج مشترك لكل المجموعات والاختبارات.

٢. الحصول على التوزيع التكراري الهامشي Marginal Frequency Distribution للدرجات في الاختبار الأول باستخدام تقديرات المعلمات على الاختبار، وتقديرات معلمات القدرة، باستخدام الدالة الرياضية الآتية:

$$\int(X) = \sum_{1-1}^{n} \int(X|\theta_g)$$

٣ . تكرار الخطوة رقم (٢) للاختبار الثاني.

٤. إجراء معادلة للاختبار بطريقة الرتب المئينية المتساوية بين الدرجات الخام في الاختبار الأول ونظيراتها في الاختبار الثاني، وذلك باستخدام التوزيع التكراري الهامشي الذي تم إنشاؤه. ومن المهم جداً في معادلة الاختبار بهذه الطريقة تغطية مدى الدرجات الخام بالكامل.

طريقة كيرنيل في المعادلة:

تتم من خلال نظام يتكون من خمس خطوات متدرجة ومتتابعة وهي:

١- مطابقة النماذج الخطية المتعددة لتوزيعات درجات الخام.

٢- تقدير الدرجات المحتملة للمجتمع المستهدف.

٣- استخدام توزيع كيرنيل لعمل اتصال بين العلامات المنفصلة في التوزيع.

٤- تطبيق المعادلة المئينية باستخدام التوزيع المتصل في الخطوة الثالثة.

٥- حساب الخطأ المعياري للمعادلة.

٨- تم إنشاء ملف يحتوي على قيم (X) وما يقابلها من قيم (Y) لكل مفحوص من المفحوصين ، حيث تم وفق هذا الملف حساب معياري تقدير دقة المعادلة وجذر متوسط مربعات الفروق.

٩- حساب الخطأ المعياري للمعادلة، وجذر متوسط مربعات الفروق.

١٠- بما أن عدد خلايا التصميم ١٨ خلية ، فقد تم إعادة تطبيق الخطوات السابقة في كل خلية من خلايا التصميم ٢٠٠ مرة وذلك للتوصل إلى تقدير دقة المعادلة ، اعتمادا على الخطأ المعياري **وجذر متوسط مربعات الفروق** لقيم المعيار الذي تم الحصول عليه عند تكرار العملية ٢٠٠ مرة. (**الحربي، ٢٠٠٩**)

الفصل الرابع

النتائج

الفصل الرابع

نتائج الدراسة

هدفت هذه الدراسة إلى مقارنة طريقة كيرنيل والمئينات وطرق نظرية استجابة الفقرة عند استخدام تصميم الفقرات المشتركة في دقة معادلة درجات الاختبارات متعددة الحدود, وذلك في ظل ظروف تجريبية مختلفة تمثلت في خلايا تضمنت المتغيرات المستقلة الآتية:

٤- حجم العينة: تم استخدام ثلاثة مستويات لحجم العينة وهي: ٢٠٠ ، ٦٠٠، ١٠٠٠ وهذه الأحجام للعينات تعتبر مناسبة لطرق المعادلة باستخدام تصميم المجموعات العشوائية.

٥- طول الاختبار وتم استخدام ثلاثة مستويات لطول الاختبار وهي ٣٠-٤٠-٦٠ فقرة.

٦- مستويات الصعوبة: وتم استخدام مستويين من الصعوبة: التشابه في الصعوبة، واختلاف الصعوبة (اختلاف الصعوبة=٠,٠١ انحراف معياري).

وقد تم التوصل الى نتائج دقة المعادلة في ظل استخدام المتغيرات السابقة منفردة ومجتمعة ، وقد تم تقدير دقة المعادلة باستخدام معياري: الخطا المعياري للمعادلة، والجذر التربيعي لمتوسط مربعات الخطا.

وفيما ياتي عرضا لنتائج الدراسة، وذلك بتناول كل متغير مستقل على حده.

<u>اولا طول الاختبار (٣٠ فقرة)</u>

<u>النتائج المتعلقة بنماذج الاختبار الذي يتألف من ٣٠ فقرة</u>

ا- التشابة في مستويات صعوبة الفقرات:

تم توليد (٣٠) فقرة اختبارية متشابهة في مستويات صعوبتها لكل نموذج (X , Y) وكذلك تم توليد (٢٠٠) عينة عشوائية من مجتمع التوزيع لكل مستوى من مستويات حجم العينة (٢٠٠ – ٦٠٠-١٠٠٠)، ثم طبقت طرق المعادلة الثلاث المستخدمة في هذه الدراسة (كيرنيل- المئينات -IRT) على هذه العينات العشوائية؛ لتقدير خطا المعادلة والجذر التربيعي لمتوسط مربعات الخطا. وفيما ياتي توضيح لنتائج كل طريقة من طرق المعادلة المستخدمة في هذه الدراسة:

<u>اولاً طريقة (IRT) في معادلة الدرجات:</u>

الخطا المعياري للمعادلة:

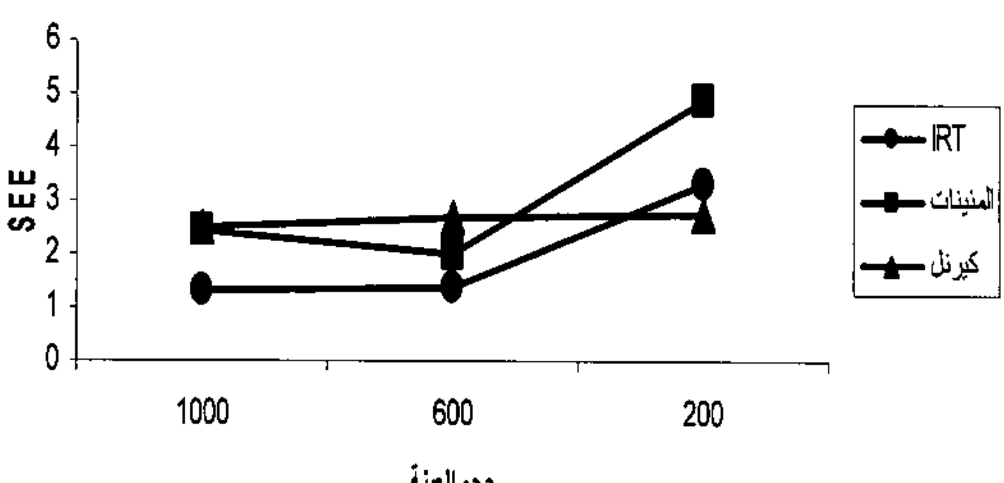
يوضح الجدول رقم (٣) قيم الخطا المعياري للمعادلة لطريقة (IRT) في معادلة الدرجات.

جدول رقم (٣)

الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٣٠) فقرة والمتشابه في الصعوبة

1000	600	200	
1.30908	1.37308	3.28443	(IRT)
2.4416	2.0257	4.90018	المئينات
2.4796	2.7017	2.71955	كيرنيل

يلاحظ من الجدول رقم (٣) ان اصغر قيمة للخطاالمعياري للمعادلة كانت عند استخدام طريقة (IRT) في معادلة الدرجات مقارنة مع طريقتي كيرنيل والمئينات، في حين كانت اكبر قيمة للخطا المعياري نتجت عن استخدام طريقة المئينات في المعادلة.كما يظهر من الجدول السابق ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة .ويوضح الشكل رقم (١) قيم الخطا المعياري للمعادلة لطريقة (IRT)على سلم الدرجات وعبر اختلاف حجم العينات.



شكل رقم (١) الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٣٠) فقرة والمتشابه الصعوبة

يوضح الشكل السابق ان طريقة(IRT) وعبر الأحجام المختلفة للعينات، كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، كما يلاحظ من الشكل السابق ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة، وان المنحنيات تميل الى الانخفاض وتقترب من الصفرعندما ترتفع حجم العينة، وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE).

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يوضح الجدول رقم (٤) قيم جذر متوسط مربعات الفروق (RMSE) لطريقة (IRT) في معادلة الدرجات.

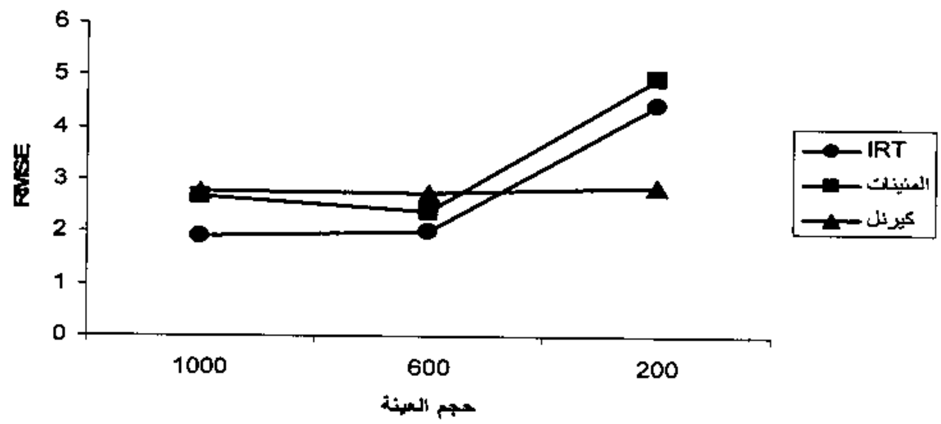
جدول رقم (٤)

قيم (RMSE) لطرق المعادلة الثلاث للاختبار الذي يتالف من ٣٠ فقرة و المتشابهة في الصعوبة

	200	600	1000
(IRT)	4.44526	2.00839	1.92461
المئينات	4.93736	2.39684	2.68813
كيرنيل	2.88424	2.75667	2.78775

يلاحظ من الجدول السابق ان اصغر قيمة من قيم (RMSE) تم الحصول عليها من استخدام طريقة (IRT) في المعادلة، في حين كانت اعلى قيمة عند استخدام طريقة المئينات في المعادلة وذلك عبر الاحجام المختلفة للعينات، كما يلاحظ من الجدول السابق ان قيمة (RMSE) تميل الى الانخفاض مع ازدياد احجام العينات.

يوضح الشكل رقم (٢) قيم (RMSE) لطريقة (IRT) في معادلة الدرجات على سلم الدرجات وعبر اختلاف احجام العينات.



شكل رقم (٢) قيم RMSE للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٣٠) فقرة والمتشابه في الصعوبة

يوضح الشكل السابق ان طريقة (IRT) وعبر الاحجام المختلفة للعينات كانت قيمة (RMSE) فيها اقل قيمة، في حين كانت اعلى قيمة عند استخدام طريقة المئينات في المعادلة وذلك عبر الاحجام المختلفة للعينات ،ويلاحظ من الشكل السابق ان قيمة (RMSE) تنخفض مع ازدياد حجم العينة.

ثانيا طريقة المئينات في معادلة الدرجات:

الخطا المعياري للمعادلة (SEE)

لنماذج الاختبار الذي يتالف من (٣٠) فقرة والمتشابهة في مستويات صعوبتها يوضح الجدول رقم (٣) نتائج قيم الخطا المعياري للمعادلة باستخدام طريقة المئينات في معادلة الدرجات وعبر الاحجام المختلفة للعينات.

ويلاحظ من الجدول رقم (٣) ان اعلى قيمة للخطا المعياري للمعادلة تم الحصول عليها من المعادلة باستخدام طريقة المئينات، كما ان قيمة الخطا المعياري للمعادلة يقل بازدياد حجم العينة.

ويوضح الشكل رقم (١) قيم الخطا المعياري للمعادلة لطريقة المئينات عبر الاحجام المختلفة للعينات.حيث يتضح من خلال هذا الشكل ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، ويلاحظ من الشكل رقم (١) ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة، كما ان المنحنيات تميل الى الانخفاض

وتقترب من الصفرعندما ترتفع حجم العينة، وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE).

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يوضح الجدول رقم (٤) قيم (RMSE) الناتجة عن تطبيق طرق المعادلة الثلاث على سلم الدرجات وعبر الأحجام المختلفة للعينات، ويلاحظ من الجدول رقم(٤) ان قيم (RMSE) تميل الى الانخفاض مع ازدياد حجم العينات، كمايوضح الشكل رقم (٢) قيم (RMSE) لطريقة المئينات عبر الاحجام المختلفة للعينات ويظهر من خلال الشكل رقم (٢) ان قيم (RMSE) تميل الى الانخفاض مع ازدياد حجم العينة.

<u>ثالثا طريقة كيرنيل في معادلة الدرجات:</u>

الخطا المعياري للمعادلة (SEE):

لنماذج الاختبار الذي يتالف من (٣٠) فقرة والمتشابه في مستويات صعوبة الفقرات يوضح الجدول رقم (٣) نتائج قيم الخطا المعياري للمعادلة باستخدام طريقة كيرنيل في معادلة الدرجات، حيث يتبين من الجدول رقم (٣) ان طريقة كيرنيل كانت افضل من طريقة المئينات و اكثر دقة، كما يلاحظ ان قيم (SEE) الناتجة عن تطبيق طرق المعادلة الثلاث تميل الى الانخفاض مع ازدياد حجم العينات.

يوضح الشكل رقم (١) قيم الخطا المعياري لطريقة كيرنيل في معادلة الدرجات عبر الاحجام المختلفة للعينات. حيث يتضح من هذا الشكل ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، كما ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة، كما يلاحظ ان المنحنيات تميل الى الانخفاض وتقترب من الصفرعندما ترتفع حجم العينة وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE).

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يظهر الجدول رقم (٤) قيم (RMSE) لطريقة كيرنيل وعبر الاحجام المختلفة للعينات، ويلاحظ من الجدول رقم (٤) ان قيم (RMSE) لطريقة كيرنيل في معادلة الدرجات كانت اقل

مقارنة مع طريقة المئينات، كما ان قيم (RMSE) لطريقة كيرنيل في المعادلة تميل الى الانخفاض مع ازدياد حجم العينة.

يوضح الشكل رقم (٢) قيم (RMSE) لطريقة كيرنيل في معادلة الدرجات.حيث يوضح الشكل رقم (٢) ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، وان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة ولطرق المعادلة الثلاث (كيرنيل–المئينات -IRT) وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE).

مقارنة نتائج المعادلة للطرق الثلاث للاختبار الذي يتالف من (٣٠) فقرة والمتشابهة في الصعوبة

فيما يتعلق بنتائج الخطأ المعياري يلاحظ من الجدول رقم (3) ان اصغر قيمة للخطا المعياري كانت عند استخدام طريقة (IRT) في معادلة الدرجات على سلم الدرجات وعبر الاحجام المختلفة للعينات ، وهذا يبين ان طريقة (IRT) هي اكثر دقة من الطرق الثلاث وعبر الاحجام المختلفة للعينات. ويظهر الجدول كذلك ان طريقة المئينات كانت اقل الطرق دقة في معادلة الدرجات وعبر الاحجام المختلفة للعينات، كما ان اختلاف قيم الخطا المعياري للمعادلة يعتمد على حجم العينة فالعينات الكبيرة تعطى دقة اكبر في معادلة الدرجات من العينات صغيرة الحجم.

كما اظهرت النتائج المتعلقة بقيم (RMSE) في الجدول رقم (4) ان طريقة (IRT) كانت اكثر دقة من الطرق الثلاث في معادلة الدرجات على الاختبارين.وتوصلت نتائج هذه الدراسة الى ان طريقة المئينات كانت اقل دقة من الطرق الثلاث في معادلة الدرجات من الاختبارين، كما يتضح من الجدول ان قيم (RMSE) تميل الى الانخفاض مع ازدياد حجم العينة.

ب- الاختلاف في مستويات الصعوبة

تم بناء نموذجين اختباريين مختلفين في مستوى صعوبة فقراتهما (X,Y) . ولكل مستوى من مستويات احجام العينات الثلاث (٢٠٠-٦٠٠-١٠٠٠)، تم توليد (٢٠٠) عينة عشوائية من مجتمع التوزيع، ثم طبقت طرق المعادلة الثلاثة المستخدمة في هذه الدراسة على هذه العينات العشوائية لتقدير دقة المعادلة ، و تم التوصل الى النتائج الاتية لكل طريقة من طرق المعادلة الثلاث:

اولا طريقة (IRT) في معادلة الدرجات:

الخطا المعياري للمعادلة:

يوضح الجدول رقم (٥) قيم الخطا المعياري لطريقة (IRT) في معادلة الدرجات

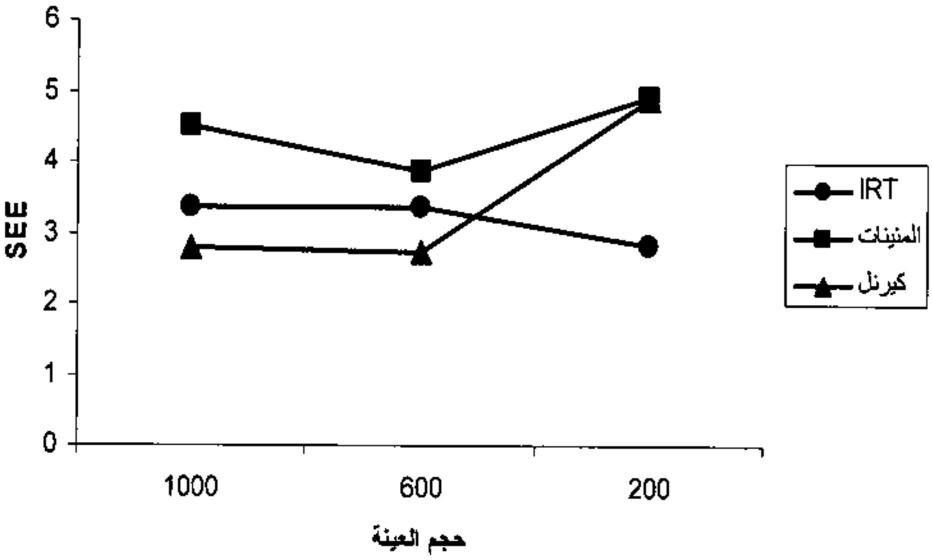
جدول رقم (٥)

الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من(٣٠) فقرة و المختلف في الصعوبة

1000	600	200	
3.37687	3.37244	2.83956	(IRT)
4.5222	3.87378	4.92831	المئينات
2.77805	2.72732	4.90018	كيرنيل

يلاحظ من الجدول السابق ان اصغر قيمة للخطا المعياري الناتج عن تطبيق طرق المعادلة الثلاث تم الحصول عليها من تطبيق طريقة كيرنيل في معادلة الدرجات، في حين بلغت اعلى قيمة للخطا المعياري من استخدام طريقة المئينات في المعادلة .

كما يلاحظ من الجدول ان قيمة الخطا المعياري للمعادلة يزداد عند ازدياد حجم العينة، وذلك عند استخدام طريقة (IRT) في حين نجده ينخفض عند استخدام طريقة كيرنيل في المعادلة .ويوضح الشكل رقم (٣) قيم الخطا المعياري للمعادلة باستخدام طريقة (IRT).



شكل رقم (٣) الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٣٠) فقرة والمختلف في الصعوبة

يوضح الشكل السابق ان طريقة كيرنيل وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، و ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة عند استخدام طريقتي كيرنيل وIRT ، وان منحنيات الطريقتين تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة، وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE)، في حين نجد ان الخطا المعياري في المعادلة لطريقة المئينات ليس له نمط محدد في الارتفاع والانخفاض.

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

ويوضح الجدول رقم (٦) قيم (RMSE) لطريقة (IRT) .

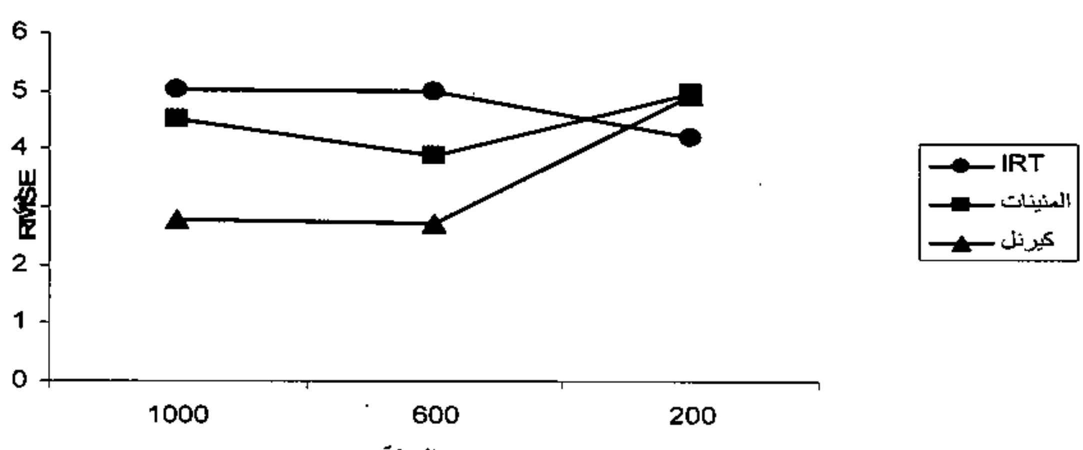
جدول رقم (٦)

قيم (RMSE) لطرق المعادلة الثلاث للاختبار الذي يتالف من (٣٠) فقرة و المختلف في مستوى الصعوبة

1000	600	200	
5.0293	4.996	4.2029	(IRT)
4.52222	3.88896	4.96286	المئينات
2.77648	2.72554	4.93736	كيرنيل

يظهر من الجدول السابق ان اصغر قيمة من قيم (RMSE) كانت عند استخدام طريقة كيرنيل في معادلة الاختبارات ، في حين بلغت اعلى قيمة من قيم (RMSE)عند استخدام طريقة (IRT) في المعادلة.كذلك دلت نتائج الجدول السابق على ان قيم (RMSE) تزداد بازدياد حجم العينة عند استخدام طريقة (IRT) في معادلة الدرجات في حين تنخفض هذه القيم عند استخدام طريقة كيرنيل .

يوضح الشكل رقم (٤) قيم (RMSE) لطريقة (IRT) في معادلة الدرجات وعبر الاحجام المختلفة للعينات .



شكل رقم (٤) قيم RMSE للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٣٠) فقرة والمختلف في الصعوبة

يوضح الشكل السابق ان طريقة كيرنيل وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، ويلاحظ كذلك ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة عند استخدام طريقتي كيرنيل وIRT.

ثانيا طريقة المئينات في معادلة الدرجات:

الخطا المعياري للمعادلة (SEE):

يوضح الجدول رقم (5) قيم الخطا المعياري لطريقة المئينات عبر الاحجام المختلفة للعينات لنماذج الاختبار التي تتالف من (٣٠) فقرة والمختلفة في مستوى صعوبة الفقرات. يلاحظ من هذا الجدول ان أعلى قيمة للخطا المعياري تم الحصول عليها من استخدام طريقة المئينات في المعادلة، في حين بلغت اقل قيمة للخطا المعياري عند استخدام طريقة كيرنيل في المعادلة.

كما يظهر الشكل رقم(٣) قيم الخطا المعياري لكل درجة على سلم الدرجات و لاحجام العينات المختلفة.حيث يظهر من الشكل رقم(٣) ان طريقة المئينات وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اعلى قيمة، ويلاحظ كذلك ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة عند استخدام طريقتي كيرنيل وIRT .

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

ويوضح الجدول رقم(٦) قيم (RMSE) لنماذج الاختبار الذي يتالف من(٣٠) فقرة و المختلفة في صعوبتها، حيث يظهر الجدول ان طريقة المئينات في المعادلة كانت اقل الطرق دقة في معادلة الدرجات وعبر الاحجام المختلفة للعينات.كما يوضح الشكل رقم(٤) قيم (RMSE)لكل درجة على سلم الدرجات عبر احجام العينات المختلفة. حيث يوضح هذا الشكل ان طريقةالمئينات وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اعلى قيمة، ويلاحظ كذلك ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة، كما ان المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة عند استخدام طريقتي كيرنيل وIRT، وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE).

ثالثا: طريقة كيرنيل

الخطا المعياري للمعادلة

يظهر الجدول رقم (٥) قيم الخطا المعياري للمعادلة (SEE) لطريقة كيرنيل على سلم الدرجات وعبر الاحجام المختلفة للعينات عندما تتالف نماذج الاختبار من(٣٠) فقرة مختلفة في صعوبتها.ويظهر الجدول ان طريقة كيرنيل كانت اكثر دقة في معادلة الدرجات مقارنة مع طريقتي (IRT) والمئينات، كما يتبين من الجدول ان قيم (SEE) تميل الى الارتفاع مع ازدياد حجم العينة وذلك عند استخدام طريقة (IRT) في معادلة الدرجات.

ويوضح الشكل رقم (٣) قيم (SEE) لطريقة كيرنيل في المعادلة وعبر الاحجام المختلفة للعينات عندما تتالف نماذج الاختبار من (٣٠) فقرة ومختلفة في صعوبتها.
يوضح الشكل رقم(٣) ان طريقة كيرنيل وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، كما ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة .

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

ويلخص الجدول رقم(٦) قيم (RMSE) لنماذج الاختبار الذي يتالف من(٣٠) فقرة و المختلفة في صعوبتها، حيث يظهر هذا الجدول ان طريقة كيرنيل في معادلة الدرجات كانت اكثر الطرق دقة في معادلة الدرجات وعبر الاحجام المختلفة للعينات.

كما يوضح الشكل(٤) قيم' (RMSE) لطريقة كيرنيل في معادلة الدرجات عند كل درجة على سلم الدرجات وعبر اختلاف احجام العينات المختلفة، حيث يلاحظ من الشكل ان قيم (RMSE) تميل الى الانخفاض مع ازدياد حجم العينة ، في حين نجد ان قيم (RMSE) تميل الى الارتفاع مع ازدياد حجم العينة.

<u>مقارنة نتائج المعادلة للطرق الثلاث للاختبار الذي يتالف من (٣٠)فقرة والمختلف في الصعوبة</u>

الخطا المعياري للمعادلة (SEE)

يستخلص من الجدول رقم (٥) ما ياتي:

- تعتبر طريقة كيرنيل من اكثر الطرق دقة في معادلة الدرجات مقارنة بطريقة المئينات وطريقة (IRT).
- تزداد قيم الخطأ المعياري للمعادلة (SEE) مع ازدياد حجم العينة وذلك عند استخدام المعادلة بطريقة المئينات (IRT)
- تقل قيم الخطأ المعياري للمعادلة (SEE) مع ازدياد حجم العينة وذلك عند استخدام طريقة كيرنيل.

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

ويستخلص من الجدول رقم(٦) ما ياتي:

١- تعتبر طريقة كيرنيل من اكثر الطرق دقة في معادلة الدرجات مقارنة مع طريقتي المئينات وطريقة (IRT).

٢- تزداد قيمة (RMSE) مع ازدياد حجم العينة وذلك عند استخدام طريقة المئينات وطريقة (IRT).

٣- تقل قيم (RMSE) مع ازدياد حجم العينة وذلك عند استخدام طريقة كيرنيل.

وفيما يلي ملخص لنتائج معادلة درجات الاختبار الذي يتكون من (٣٠) فقرة

جدول رقم (٧)

ملخص نتائج طرق المعادلة الثلاثة لنماذج الاختبار الذي يتكون من (٣٠) فقرة

طريقة المعادلة	التشابه في الصعوبة						الاختلاف في الصعوبة					
	200		600		1000		200		600		1000	
	SE	RMSE	SE	RMSE	SE	RMSE	SE	RMSE	SE	RMSE	SE	RMSE
IRT	١	١	٢	٣	٢	٣	٢	٢	١	١	١	١
المئينات	٣	٣	٣	٢	٣	٢	٣	٣	٢	٢	٢	٢
كيرنيل	٢	٢	١	١	١	١	١	١	٣	٣	٣	٣

ثانيا: طول نماذج الاختبار (٤٠) فقرة

أ- نماذج الاختبارالمتشابهة في مستوى الصعوبة

تم توليد نموذجين للاختبار من (٤٠) فقرة متشابهين في مستوى الصعوبة (X,Y) بعد ذلك تم توليد العينات العشوائية وباحجام (٢٠٠-٦٠٠-١٠٠٠) من مجتمع التوزيع طبقت طرق المعادلة الثلاثة المستخدمة في هذه الدراسة (كيرنيل-المئينات-(IRT)) على هذه العينات العشوائية لتقدير خطا المعادلة وجذر متوسط مربعات الفروق، وفي ما ياتي توضيح لنتائج كل طريقة من طرق المعادلة المستخدمة .

اولا طريقة (IRT) في معادلة الدرجات

الخطا المعياري للمعادلة

يوضح الجدول رقم (٨) قيم الخطا المعياري للمعادلة باستخدام طريقة (IRT) في معادلة الدرجات وعبر الاحجام المختلفة للعينات.

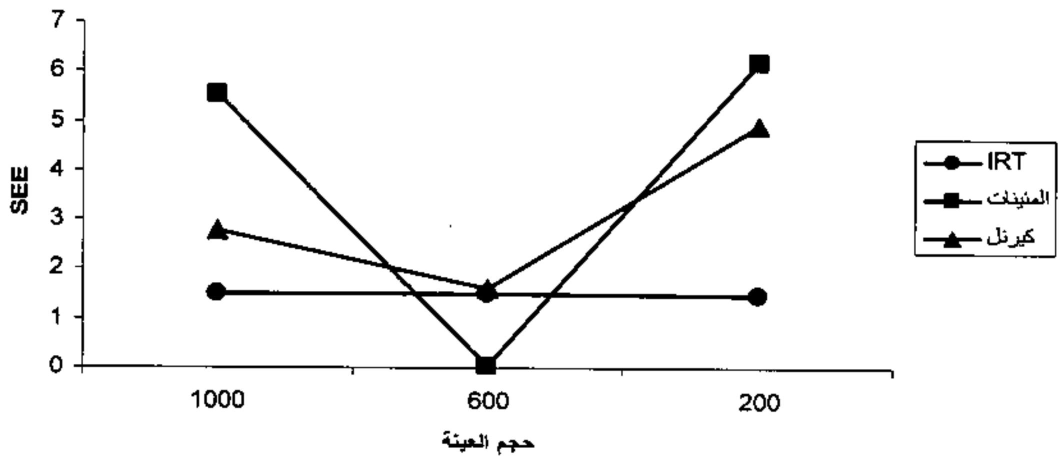
جدول رقم (٨)

الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من ٤٠ فقرة و المتشابه في الصعوبة

	200	600	1000
(IRT)	1.46867	1.51999	1.49027
المئينات	6.1899	0.0449	5.5182
كيرنيل	4.90018	1.6004	2.77805

يظهر من الجدول السابق ان اقل قيمة للخطا المعياري للمعادلة (SEE) تم الحصول عليها عند استخدام طريقة المئينات وذلك عندما كان حجم العينة (٦٠٠)، في حين بلغت اعلى قيمة للخطا المعياري عند استخدام طريقة المئينات في المعادلة وعندما كان حجم العينة (٢٠٠). كما ويبين الجدول ان طريقة (IRT) كانت اكثر دقة في معادلة الدرجات مقارنة مع طريقتي المئينات و كيرنيل، وان قيمة الخطا المعياري للمعادلة تنخفض عند زيادة حجم العنية.

يوضح الشكل رقم (٥) قيم الخطا المعياري للمعادلة باستخدام طريقة (IRT) على سلم الدرجات وعبر الأحجام المختلفة للعينات (٢٠٠-٦٠٠-١٠٠٠) عندما تتالف نماذج الاختبار من (٤٠) فقرة ومتشابهة في صعوبتها.



شكل رقم(٥) الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٤٠) فقرة والمتشابه في الصعوبة.

ويظهر من الشكل السابق ان اقل قيمة للخطا المعياري للمعادلة (SEE) تم الحصول عليها عند استخدام طريقة المئينات وذلك عندما كان حجم العينة (٦٠٠)، في حين بلغت اعلى قيمة للخطا المعياري عند استخدام طريقة المئينات في المعادلة عندما كان حجم العينة (٢٠٠).

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

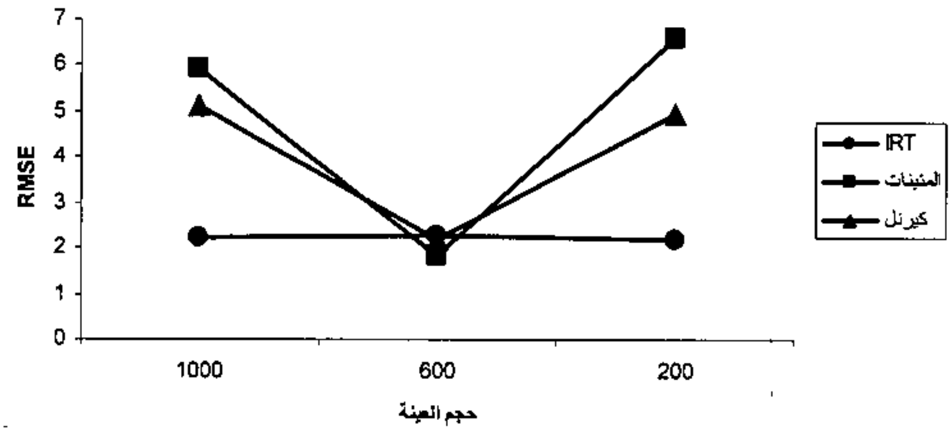
ولنماذج الاختبار الذي يتألف من ٤٠ فقرة يوضح الجدول رقم (٩) قيم (RMSE)لطرق المعادلة الثلاث

جدول رقم (٩)

قيم (RMSE) لطرق المعادلة الثلاث للاختبار الذي يتالف من ٤٠ فقرة و المتشابهة في الصعوبة

1000	600	200	
2.22096	2.26779	2.1688	(IRT)
5.91133	1.83657	6.5815	المئينات
5.13248	2.16753	4.93736	كيرنيل

يلاحظ من الجدول السابق ان اقل قيمة من قيم (RMSE) تم الحصول عليها عند استخدام طريقة المئينات في معادلة الدرجات وعند استخدام حجنم عينة (٦٠٠)، في حين بلغت اعلى قيمة من قيم (RMSE)عند استخدام طريقة المئينات و ذلك عند استخدام حجم عينة (٢٠٠) كذلك يلاحظ من الجدول عدم وجود علاقة بين حجم العينة و قيم (RMSE)،ويوضح الشكل رقم (٦) قيم (RMSE)باستخدام طريقة (IRT)على سلم الدرجات وعبر الاحجام المختلفة (٢٠٠-٦٠٠-١٠٠٠)



شكل رقم (٦) قيم RMSE للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٤٠) فقرة والمتشابه في الصعوبة

يلاحظ من الشكل السابق ان اقل قيمة من قيم (RMSE) تم الحصول عليها عند استخدام طريقة المئينات وذلك عند استخدام حجم عينة (٦٠٠) ، في حين بلغت اعلى قيمة من قيم (RMSE) عند استخدام طريقة المئينات في معادلة الدرجات و ذلك عند استخدام حجم عينة (٢٠٠).كذلك يلاحظ من الجدول عدم وجود علاقة بين حجم العينة و قيم (RMSE).

ثانيا طريقة المئينات في معادلة الدرجات:

الخطأ المعياري للمعادلة (SEE):

يوضح الجدول رقم (٨) قيم الخطأ المعياري للمعادلة باستخدام طريقة المئينات لنماذج الاختبار الذي يتكون من (٤٠) فقرة و المتشابهة في صعوبتها وعبر الأحجام المختلفة للعينات. ويلاحظ من الجدول رقم (٨) ان طريقة(IRT) كانت اقل الطرق دقة في معادلة الدرجات كما يبين الجدول ان قيمة الخطأ المعياري يزداد اذا قلت حجم العينة في حين نجده ينخفض اذا زادت حجم العينة.

ويظهر الشكل رقم (٥) قيم الخطأ المعياري للمعادلة لكل درجة على سلم الدرجات عبر اختلاف احجام العينات. حيث يوضح هذا الشكل ان طريقة(المئينات) وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، كما ان قيمة الخطا المعياري للمعادلة لا يتاثر مع ازدياد حجم العينة او نقصانها.

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

لنماذج الاختبار الذي يتألف من (٤٠) فقرة يوضح الجدول رقم (٩) قيم (RMSE) لطرق المعادلة الثلاث، حيث يلاحظ من الجدول رقم (٩) ان طريقة(IRT) كانت اكثر الطرق دقة في معادلة الدرجات, تلتها طريقة كيرنيل, في حين احتلت طريقة المئينات المرتبة الثالثة من حيث دقة المعادلة مقارنة بالطرق الاخرى، ويظهر الجدول رقم (٨) ان قيم (RMSE) لا تتاثر مع ازدياد حجم العينة او نقصانها.

و يوضح الشكل رقم (٦) قيم (RMSE) لطريقة المئينات و عبرالاحجام المختلفة للعينات (٢٠٠-٦٠٠-١٠٠٠).حيث يلاحظ من الشكل رقم (٦) ان طريقةIRT كانت اكثر الطرق دقة في معادلة الدرجات, تلتها طريقة كيرنيل, في حين احتلت طريقة المئينات المرتبة الثالثة من حيث دقة المعادلة مقارنة بالطرق الاخرى.

ثالثا طريقة كيرنيل في معادلة الدرجات

الخطأ المعياري للمعادلة (SEE)

لنماذج الاختبار الذي يتالف من (٤٠) فقرة والمتشابه في مستوى الصعوبة، يوضح الجدول رقم (٨) نتائج قيم الخطأ المعياري للمعادلة (SEE) باستخدام طريقة كيرنيل. ويظهر من الجدول ان طريقة كيرنيل مقارنة بطريقة المئينات كانت اكثر دقة في معادلة الدرجات على سلم الدرجات وعبر الاحجام المختلفة للعينات،.ويظهر الجدول كذلك انه لا يوجد علاقة بين حجم العينة والخطا المعياري في المعادلة.

ويوضح الشكل رقم (٥) قيم الخطا المعياري لطريقة كيرنيل في معادلة الدرجات على سلم الدرجات وعبر الاحجام المختلفة للعينات.حيث يوضح هذا الشكل ان طريقةالمئينات وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، ويظهر الشكل كذلك انه لا يوجد علاقة بين حجم العينة والخطا المعياري في المعادلة.

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

لنماذج الاختبار الذي يتألف من (٤٠) فقرة يوضح الجدول رقم (٩) قيم (RMSE) لطرق المعادلة الثلاث، حيث يبين الجدول ان طريقة (IRT) كانت اكثر الطرق دقة في معادلة الدرجات, تلتها طريقة كيرنيل, في حين احتلت طريقة المئينات المرتبة الثالثة من حيث دقة المعادلة مقارنة بالطرق الاخرى.

ويوضح الشكل رقم (٦) قيم (RMSE) لطريقة كيرنيل في معادلة الدرجات وعبر الأحجام المختلفة للعينات (٢٠٠-٦٠٠-١٠٠٠).

يوضح الشكل رقم(٦) ان طريقةالمئينات وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اعلى قيمة، ويظهر هذا الشكل ان قيمة الخطا المعياري للمعادلة لا تتاثر بحجم العينةزيادة اونقصانا.

مقارنة نتائج المعادلة للطرق الثلاث للاختبار الذي يتالف من (٤٠)فقرة والمتشابهة في الصعوبة

النتائج المتعلقة بالخطا المعياري للمعادلة

يلاحظ من الجدول رقم (٨) ان طريقة (IRT) كانت اكثر الطرق دقة في معادلة الدرجات, تلتها طريقة كيرنيل,ثم المئينات كما يلاحظ من الجدول ان طريقة المئينات كانت اقل الطرق دقة في معادلة الدرجات عبر الاحجام المختلفة للعينات كما ان اختلاف حجم العينات لا يؤثر في قيمة الخطا المعياري للمعادلة.

النتائج المتعلقة بقيم الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يلاحظ من الجدول رقم (٩) ان طريقة (IRT) كانت اكثر الطرق دقة من الطرق الثلاث في معادلة الدرجات. وتوصلت نتائج هذه الدراسة ان قيم (RMSE) الناتجة عن استخدام طريقة المئينات في المعادلة كانت اقل الطرق دقة من الطرق الاخرى في معادلة الدرجات، كما يلاحظ انه لا يوجد علاقة بين قيم (RMSE) وحجم العينة.

ب- نماذج الاختبار المختلفة في مستوى الصعوبة

تم بناء نمدوجين اختباريين مكونين من (٤٠) فقرة مختلفين في مستوى الصعوبة(Y,X) لكل مستوى من مستويات حجم العينة, حيث ولدت (٢٠٠) عينة عشوائية من مجمتع التوزيع. ثم طبقت طرق المعادلة الثلاث المستخدمة في هذه الدراسة على هذه العينات العشوائية لتقدير دقة المعادلة,وتم التوصل الى النتائج الاتية لكل طريقة من طرق المعادلة الثلاثة :

اولا طريقة (IRT) في معادلة الدرجات

الخطا المعياري للمعادلة

يوضح الجدول رقم (١٠) قيم الخطا المعياري لطريقة (IRT)

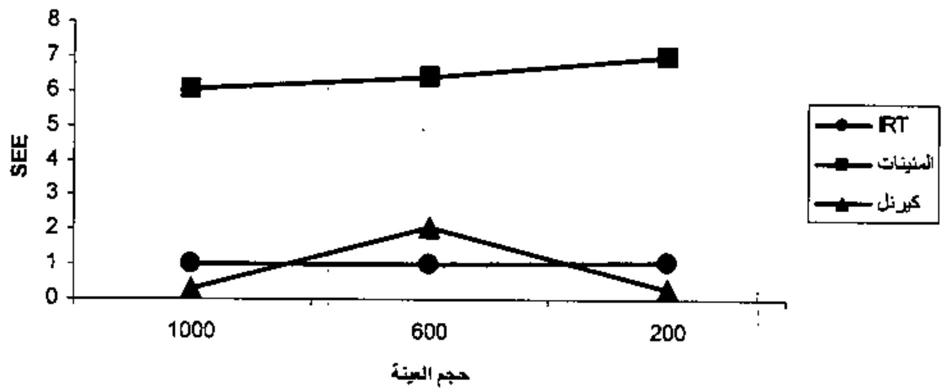
جدول رقم (١٠)

الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٤٠) فقرة والمختلف في الصعوبة

	200	600	1000
(IRT)	1.04575	1.01326	1.0193
المئينات	7.0177	6.4219	6.0543
كيرنيل	0.2959	2.0592	0.2693

يلاحظ من الجدول السابق ان اصغر قيمة للخطا المعياري تم الحصول عليها عند استخدام طريقة كيرنيل, في حين بلغت اعلى قيمة للخطأ المعياري للمعادلة باستخدام طريقة المئينات، كما يلاحظ من الجدول السابق ان قيمة الخطا المعياري للمعادلة يقل عند ازدياد حجم العينة.

ويوضح الشكل رقم (٧) قيم الخطأ المعياري للمعادله باستخدام طريقه (IRT) .



شكل رقم (٧) الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٤٠) فقرة والمختلف في الصعوبة

يوضح الشكل السابق ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، ويلاحظ من الشكل السابق ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة، كما يلاحظ ان المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة، وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE)

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

لنماذج الاختبار الذي تتكون من (٤٠) فقره والمختلفة في صعوبتها يوضح الجدول رقم(١١) قيم (RMSE) لطريقة (IRT) في المعادلة

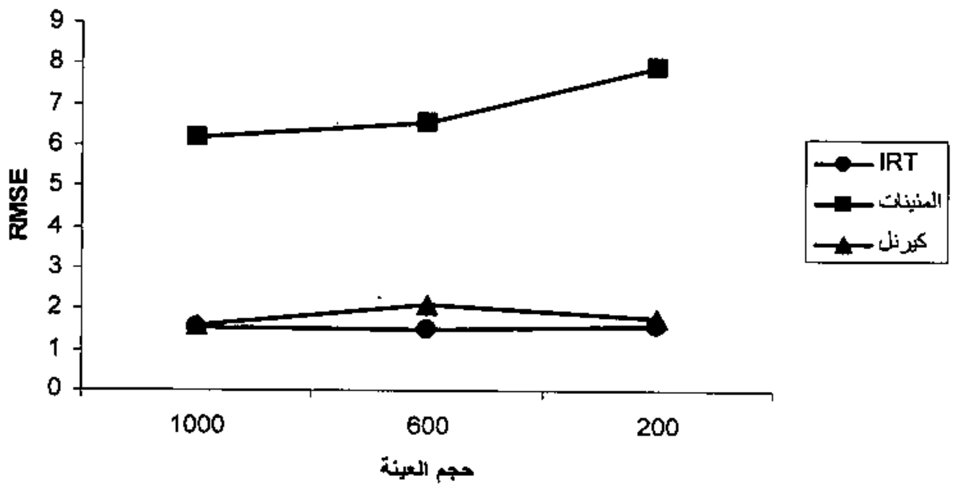
جدول رقم(١١)

قيم (RMSE) لطرق المعادلة الثلاث للاختبار الذي يتالف من ٤٠ فقرة و المختلف في الصعوبة

	200	600	1000
(IRT)	1.57426	1.50813	1.53039
المئينات	7.9113	6.5411	6.1891
كيرنيل	1.79073	2.09861	1.57167

يلاحظ من الجدول السابق ان طريقة (IRT) كانت اكثر الطرق دقة في معادلة الدرجات, تلتها طريقة كيرنيل. كما يلاحظ ان طريقة المئينات كانت اقل الطرق دقة في دقة معادلة الدرجات

وعبر الاحجام المختلفة للعينات. ويظهر الجدول قيم RMSE تتأثر بحجم العينات, فكلما زادت حجم العينة قلت قيم RMSE, واذا انخفضت حجم العينة ازداد ا قيم RMSE وعبر الاحجام المختلفة للعينات. ويوضح الشكل رقم (٨) قيم (RMSE) لطريقة (IRT)



شكل رقم (٨) قيم RMSE للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٤٠) فقرة والمختلفة في الصعوبة

يوضح الشكل السابق ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيم RMSE للمعادلة فيها اقل قيمة ، وان قيم RMSE للمعادلة تنخفض مع ازدياد حجم العينة. كما يلاحظ ان المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة، وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيم RMSE.

ثانيا: طريقة المئينات في معادلة الدرجات

النتائج المتعلقة بالخطا المعياري للمعادلة

يوضح الجدول رقم (١٠) قيم الخطا المعياري لطريقة المئينات وعبر الاحجام المختلفة للعينات لنماذج الاختبار الذي يتالف من (٤٠) فقرة والمختلف في مستوى الصعوبة.وتظهر نتائج هذا الجدول ان طريقة المئينات كانت اقل الطرق دقة في معادلة درجات النماذج. و ان قيم الخطا المعياري في المعادلة المئينية تقل بازدياد حجم العينة.

ويظهر الشكل رقم (٧) قيم الخطا المعياري للمعادلة المئينية ولكل درجة على سلم الدرجات وعبر الاحجام المختلفة للعينات.حيث يوضح هذا الشكل ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة ، وان قيمة الخطا المعياري

للمعادلة تنخفض مع ازدياد حجم العينة. كما يلاحظ ان المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة، وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE).

النتائج المتعلقة بقيم الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يوضح الجدول رقم (١١) قيم (RMSE) لطريقة المئينات وعبر الاحجام المختلفة للعينات لنماذج الاختبار الذي يتالف من (٤٠) فقرة والمختلف في مستوى الصعوبة .واظهرت نتائج الجدول السابق ان طريقة المئينات في المعادلة كانت اقل الطرق دقة في معادلة درجات النماذج.كما يلاحظ ان قيم (RMSE) في المعادلة المئينية تقل بازدياد حجم العينة.

ويظهر الشكل رقم (٨) قيم (RMSE) للمعادلة المئينية ولكل درجة على سلم الدرجات وعبر الاحجام المختلفة للعينات.

يوضح الشكل السابق ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيم RMSE فيها اقل قيمة ويلاحظ من الشكل السابق ان قيم RMSE تنخفض مع ازدياد حجم العينة .

ثالثا: طريقة كيرنيل في معادلة الدرجات :

النتائج المتعلقة بالخطا المعياري للمعادلة

يوضح الجدول رقم (١٠) قيم الخطا المعياري للمعادلة باستخدام طريقة كيرنيل للاحجام المختلفة للعينات لنماذج الاختبار الذي يتالف من (٤٠) فقرة والمختلف في مستوى الصعوبة، حيث أظهرت نتائج هذا الجدول ان طريقة(IRT) كانت اكثر الطرق دقة مقارنة مع طريقة المئينات وطريقة كيرنيل في معادلة الدرجات، كما يلاحظ ان قيم الخطا المعياري في المعادلة المئينية تقل بازدياد حجم العينة.

ويظهر الشكل رقم (٧) قيم الخطا المعياري لطريقة كيرنيل في معادلة الدرجات وعبر الاحجام المختلفة للعينات. يوضح الشكل رقم (٧) ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، كما يلاحظ ان المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE).

النتائج المتعلقة بقيم الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

لنماذج الاختبار الذي يتألف من (٤٠) فقرة يوضح الجدول رقم (١١) قيم (RMSE) لطرق معادلة الثلاث ويلاحظ من الجدول طريقة(IRT) كانت اكثر الطرق دقة في معادلة الدرجات , تلتها طريقة كيرنيل , في حين احتلت طريقة المئينات المرتبة الثالثة من حيث دقة المعادلة مقارنة بالطرق الاخرى. ويظهر الجدول رقم (١١) ان قيم (RMSE) لطريقة كيرنيل تقل بازدياد حجم العينة والعكس صحيح.

ويوضح الشكل رقم (٨) قيم (RMSE) لطريقة كيرنيل في معادلة الدرجات على سلم الدرجات وعبر الاحجام المختلفة للعينات (٢٠٠-٦٠٠-١٠٠٠).

يوضح الشكل السابق ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيم (RMSE) فيها اقل قيمة ، كمايلاحظ من الشكل رقم (٨) ان قيم (RMSE) تنخفض مع ازدياد حجم العينة كما يلاحظ ان المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيم (RMSE).

مقارنة نتائج المعادلة للطرق الثلاث للاختبار الذي يتالف من (٤٠) فقرة والمختلف في الصعوبة

النتائج المتعلقة بالخطا المعياري للمعادلة

يلاحظ من الجدول رقم(١٠) ان طريقة (IRT) من اكثر الطرق دقة في معادلة الدرجات من طريقة المئينات و كيرنيل, كما اظهرت النتائج ان قيم الخطا المعياري للمعادلة تقل مع ازدياد حجم العينة وتزداد مع نقصان حجم العينة.

النتائج المتعلقة بقيم الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يلاحظ من الجدول رقم (١١) ان طريقة (IRT) كانت اكثر الطرق دقة في معادلة درجات الاختبار على سلم الدرجات وعبر الاحجام المختلفة للعينات, بينما كانت طريقة المئينات اقل الطرق دقة في معادلة الدرجات.كما يظهر الجدول السابق ان قيم (RMSE) تتاثر بحجم العينات فكلما زادت حجم العينات قلت قيم (RMSE),واذا قلت حجم العينة تزداد قيمة (RMSE).

وفيما يلي ملخص لنتائج معادلة درجات الاختبار الذي يتكون من (٤٠) فقرة

جدول رقم (١٢)

ملخص نتائج طرق المعادلة الثلاثة لنماذج الاختبار الذي يتكون من (٤٠) فقرة

طريقة المعادلة	التشابه في الصعوبة						الاختلاف في الصعوبة					
	200		600		1000		200		600		1000	
	SE	RMSE	SE	RMSE	SE	RMSE	SE	RMSE	SE	RMSE	SE	RMSE
IRT	١	١	٣	٣	١	١	٢	١	١	١	٢	١
المئينات	٣	٣	١	١	٣	٣	٣	٣	٣	٣	٣	٣
كيرنيل	٢	٢	٢	٢	٢	٢	١	٢	٢	٢	١	٢

ثالثا: طول الاختبار (٦٠) فقرة

تم توليد نموذجين للاختبار حيث اشتمل كل اختبار على (٦٠) فقرة من نوع الاختبار متعدد ، بعد ذلك تم توليد(٢٠٠) عينة عشوائية و بمستويات (٢٠٠-٦٠٠-١٠٠٠) لكل عينة من مجتمع التوزيع.

أ - التشابه في مستويات صعوبة النماذج

تم توليد نموذجين للاختبار بحيث اشتمل كل اختبار على (٦٠) فقرة متشابه في صعوبته، ثم ولدت (٢٠٠) عينة عشوائية و بمستويات (٢٠٠-٦٠٠-١٠٠٠) لكل عينة من مجتمع التوزيع، طبقت طرق المعادلة الثلاث المستخدمة في هذه الدراسة (IRT-المئينات-كيرنيل) لتقدير خطأ المعادلة وجذر متوسطات مربعات الفروق وفيما ياتي توضيح لنتائج كل طريقة من طرق المعادلة المستخدمة:

اولا طريقة (IRT) في معادلة الدرجات

الخطا المعياري للمعادله(SEE)

يوضح الجدول رقم(١٣) قيم الخطا المعياري للمعادله باستخدام طريقة(IRT)

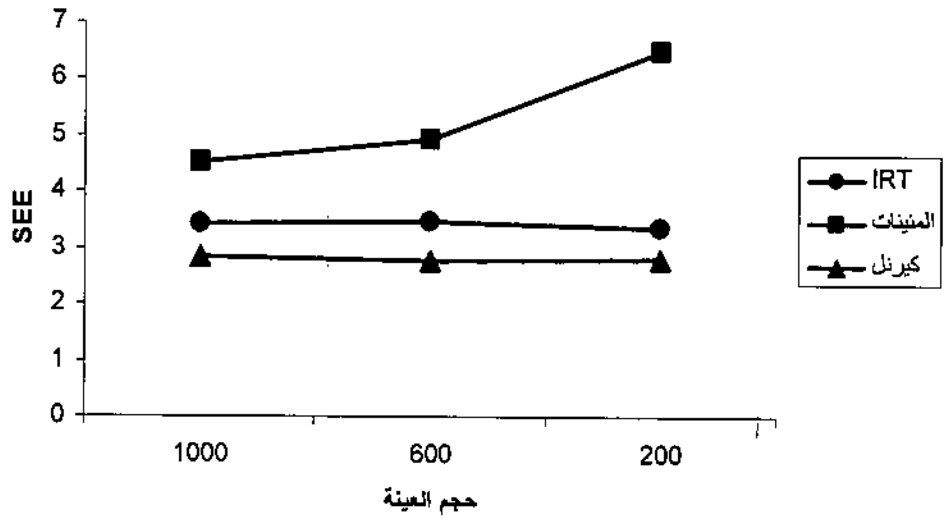
جدول رقم(١٣)

الخطا المعياري للمعادله لطرق المعادلة الثلاث للاختبار الذي يتالف من (٦٠) فقره و المتشابه في الصعوبة

	200	600	1000
(IRT)	3.36806	3.47922	3.42516
المئينات	6.48626	4.91678	4.53104
كيرنيل	2.79401	2.76196	2.83025

يظهر من الجدول السابق ان اقل قيمه للخطا المعياري للمعادله كانت عند استخدام طريقة كيرنيل في المعادلة، في حين كانت اكبر قيمه للخطا المعياري نتجن عن استخدام طريقة المئينات في معادلة الدرجات، ويلاحظ من الجدول السابق ان قيم الخطأ المعياري تقل بازدياد حجم العينه والعكس صحيح.

ويظهر الشكل رقم (٩) قيم الخطا المعياري لطريقة (IRT)على سلم الدرجات و عبر الاحجام المختلفه للعينات (٢٠٠-٦٠٠ - ١٠٠٠).



شكل رقم (٩) الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٦٠) فقرة والمتشابه في الصعوبة

يوضح الشكل السابق ان طريقة كيرنيل وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، ويلاحظ من الشكل السابق ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة .

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

لنماذج الاختبار الذي يتالف من (٦٠) فقره يوضح الجدول رقم(١٤) قيم (RMSE) لطرق المعادلة الثلاثة

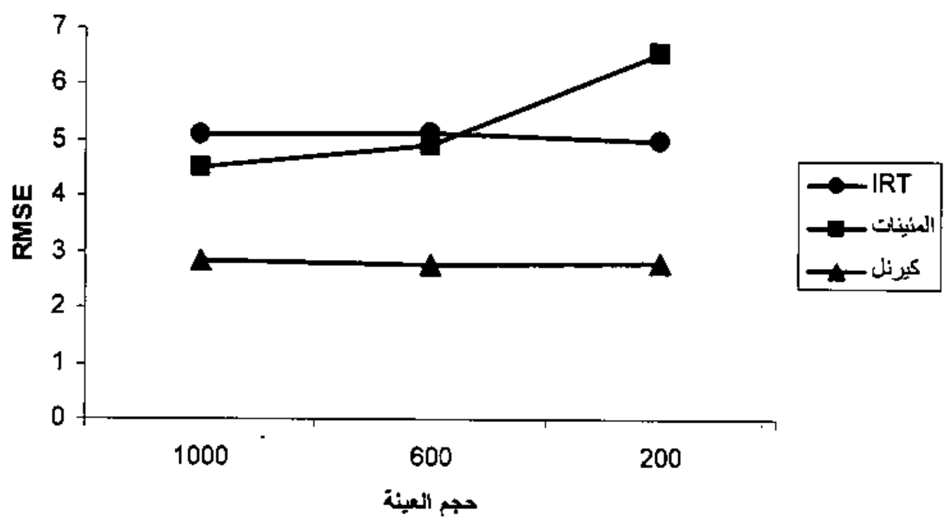
جدول رقم(١٤)

قيم (RMSE) لطرق المعادلة الثلاث للاختبار الذي يتالف من (٦٠) فقرة و المتشابهة في الصعوبة

1000	600	200	
5.10337	5.12576	4.99398	(IRT)
4.52222	4.90884	6.57401	المئينات
2.82917	2.76105	2.79262	كيرنيل

يظهر الجدول السابق ان طريقة كيرنيل كانت اكثر الطرق دقه في معادلة الدرجات وعبر الاحجام المختلفه للعينات، بينما كانت طريقة المئينات اقل الطرق دقه في معادلة الدرجات وعبر الاحجام المختلفه للعينات، ويظهر الجدول ان قيم (RMSE) تتاثر بحجم العينه فكلما كبر حجم العينه قلة قيمة (RMSE) واذا نقص حجم العينه ارتفعت قيم (RMSE).

يوضح الشكل رقم(١٠) قيم (RMSE)عند استخدام طريقة (IRT) في معادلة الدرجات على سلم الدرجات وعبر الاحجام المختلفه للعينات (٢٠٠-٦٠٠-١٠٠٠)



شكل رقم (١٠) قيم RMSE للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٦٠) فقرة والمتشابه في الصعوبة

يظهر الشكل رقم (١٠) ان طريقة كيرنيل كانت اكثر الطرق دقه في معادلة الدرجات وعبرالاحجام المختلفه للعينات، بينما كانت طريقة المئينات اقل الطرق دقه في معادلة الدرجات وعبر الاحجام المختلفه للعينات، ويظهر الجدول ان قيم (RMSE) تتاثر بحجم العينه فكلما كبر حجم العينه قلة قيمة (RMSE) واذا نقص حجم العينه ارتفعت قيم (RMSE).

ثانيا طريقة المئينات في معادلة الدرجات

الخطا المعياري للمعادلة (SEE)

يوضح الجدول رقم (١٣) قيم الخطا المعياري للمعادلة باستخدام طريقة المئينات في معادلة الدرجات، حيث كانت طريقة المئينات من اقل الطرق دقة في معادلة الدرجات مقارنة مع

طريقة (IRT) وطريقة كيرنيل، كما يظهر الجدول وجود علاقة ما بين حجم العينة وقيمة الخطا المعياري حيث تتاثر قيم الخطا المعياري بحجم العينة زيادة او نقصانا.

ويظهر الشكل رقم (٩) قيم الخطأ المعياري لطريقة المئينات على سلم الدرجات عبر الاحجام المختلفة للعينات.حيث يتضح من الشكل رقم (٩) ان طريقة كيرنيل وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، في حين بلغت اعلى قيمة للخطا المعياري عند استخدام الطريقة المئينية في المعادلة، ويلاحظ من الشكل السابق ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة .

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يوضح الجدول رقم (١٤) ان طريقة كيرنيل كانت اكثر الطرق دقة في معادلة الدرجات و عبر الاحجام المختلفة للعينات، بينما كانت طريقة المئينات اقل الطرق دقة في معادلة الدرجات وعبر الاحجام المختلفة للعينات، كما يظهر الجدول ان قيم (RMSE) تنخفض بازدياد حجم العينة و ترتفع بانخفاض جحم العينة.

يوضح الشكل رقم (١٠) قيم (RMSE) باستخدام طريقة المئينات على سلم الدرجات و عبر الاحجام المختلفة للعينات

يوضح الشكل السابق ان طريقة كيرنيل كانت اكثر الطرق دقة في معادلة الدرجات و عبر الاحجام المختلفة للعينات، بينما كانت طريقة المئينات اقل الطرق دقة في معادلة الدرجات وعبر الاحجام المختلفة للعينات، كما يظهر الجدول ان قيم (RMSE) تنخفض بازدياد حجم العينة و ترتفع بانخفاض حجم العينة.

ثالثا طريقة كيرنيل في معادلة الدرجات

الخطا المعياري للمعادلة (SEE)

لنماذج الاختبار الذي يتالف من (٦٠) فقرة و المتشابهة في الصعوبة، يوضح الجدول رقم (١٣) نتائج الخطأ المعياري للمعادلة باستخدام طريقة كيرنيل. ويظهر الجدول رقم (١٣) ان طريقة كيرنيل كانت اكثر الطرق دقة في معادلة الاختبارات مقارنة بطريق مئينات (IRT)، ويظهر الجدول كذلك انه كلما ارتفعت حجم العينة قلت قيم (RMSE) وكلما قل حجم العينة ارتفعت قيم (RMSE).

ويوضح الشكل رقم (٩) قيم الخطأ المعياري لطريقة كيرنيل في معادلة الدرجات وعبر الاحجام المختلفة للعينات ، حيث يوضح الشكل السابق ان طريقة كيرنيل وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة كما يلاحظ من هذا الشكل ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة، كما ان المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE).

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يظهر الجدول رقم (١٤) ان طريقة كيرنيل كانت اكثر الطرق دقه في معادلة الاختبارات عبر الاحجام المختلفة للعينات، ويظهر الجدول ان قيم (RMSE) تتاثر بحجم العينة فاذا زادت حجم العينة قلت قيم (RMSE) واذا قلت حجم العينة ارتفعت قيم (RMSE).

يوضح الشكل رقم (١٠) قيم (RMSE) باستخدام طريقة كيرنيل على سلم الدرجات عبر الاحجام المختلفة للعينات،.حيث يوضح هذا الشكل ان طريقة كيرنيل وعبر الأحجام المختلفة للعينات كانت قيم (RMSE) فيها اقل قيمة، كماان قيم (RMSE) تنخفض مع ازدياد حجم العينة، كما يلاحظ ان المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيم (RMSE).

مقارنة نتائج المعادلة للطرق الثلاث للاختبار الذي يتالف من (٦٠) فقرة والمتشابهة في الصعوبة

النتائج المتعلقة بالخطأ المعياري للمعادلة

يلاحظ من الجدول رقم (١٣) ان طريقة كيرنيل كانت من اكثر الطرق دقة في معادلة الدرجات وعبر الاحجام المختلفة للعينات, تلتها طريقة (IRT) في معادلة الدرجات, ثم المئينات. كما تعتبر طريقة المئينات من اقل الطرق دقة مقارنة مع طريقة (IRT), وكيرنيل وعبر الاحجام المختلفة للعينات، وتظهر نتائج الجدول رقم (١٣) ان قيم الخطأ المعياري للمعادلة تميل الى الارتفاع مع نقصان حجم العينة, ويميل الى الانخفاض مع ارتفاع حجم العينة.

النتائج المتعلقة بقيم الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يلاحظ من الجدول رقم (١٤) ان طريقة كيرنيل كانت من اكثر الطرق دقة في معادلة الدرجات وعبر الاحجام المختلفة للعينات.كما اظهرت النتائج ان قيم (RMSE) تتاثر بحجم العينة, فاذا زادت حجم العينة قلت قيم (RMSE) واذا قلت حجم العينة زادت قيم (RMSE) .

ب- الاختلاف في مستويات الصعوبة .

تم بناء نمودجين للاختبار مكونين من (٦٠) فقرة ومختلفين في مستوى الصعوبة, ثم تم توليد (٢٠٠) عينة عشوائية ولكل مستوى من مستويات حجم العينة(٢٠٠-٦٠٠-١٠٠٠). طبقت طرق المعادلة الثلاث المستخدمة في هذه الدراسة على هذه العينات العشوائية لتقدير الخطا المعياري للمعادلة وجذر متوسط مربعات الفروق في المعادلة, تم التوصل الى النتائج الاتية لكل طريقة من طرق المعادلة الثلاثة:

اولا:طريقة (IRT) في معادلة الدرجات:

الخطا المعياري للمعادلة (SEE):

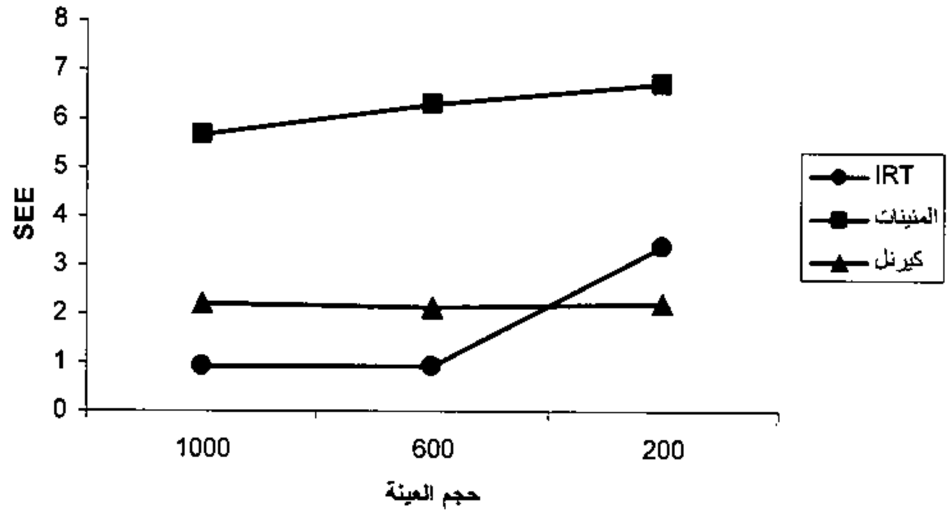
يوضح الجدول رقم (١٥) قيم الخطا المعياري للمعادلة لطريقة (IRT) في معادلة درجات الاختبار وعبر الاحجام المختلفة للعينات.

جدول رقم (١٥)

الخطا المعياري للمعادلة للطرق الثلاث للاختبار الذي يتالف من ٦٠ فقرة والمختلف في الصعوبة

1000	600	200	
0.90062	0.91555	3.38358	(IRT)
5.64892	6.30457	6.7232	المئينات
2.21651	2.10637	2.20815	كيرنيل

يلاحظ من الجدول السابق ان اقل قيمة للخطا المعياري تم الحصول عليها عند استخدام طريقة (IRT) , في حين بلغت اعلى قيمة للخطا المعياري عند استخدام طريقة المئينات. كما يلاحظ من الجدول السابق ان قيمة الخطا المعياري للمعادلة يقل عند ازدياد حجم العينة, في حين تزداد قيمة الخطا المعياري للمعادلة عندما تقل حجم العينة.ويوضح الشكل رقم (١١)المعادلة باستخدام طريقة IRT



شكل رقم (١١) الخطا المعياري للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٦٠) فقرة والمختلف في الصعوبة

يوضح الشكل السابق ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، كما ان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة، وان المنحنيات تميل الى النزول وتقترب من الصفرعندما يرتفع حجم العينة، وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيمة الخطا المعياري للمعادلة (SEE).

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يوضح الجدول رقم (١٦) قيم (RMSE) لطريقة (IRT) في معادلة الدرجات وعبر الاحجام المختلفة للعينات.

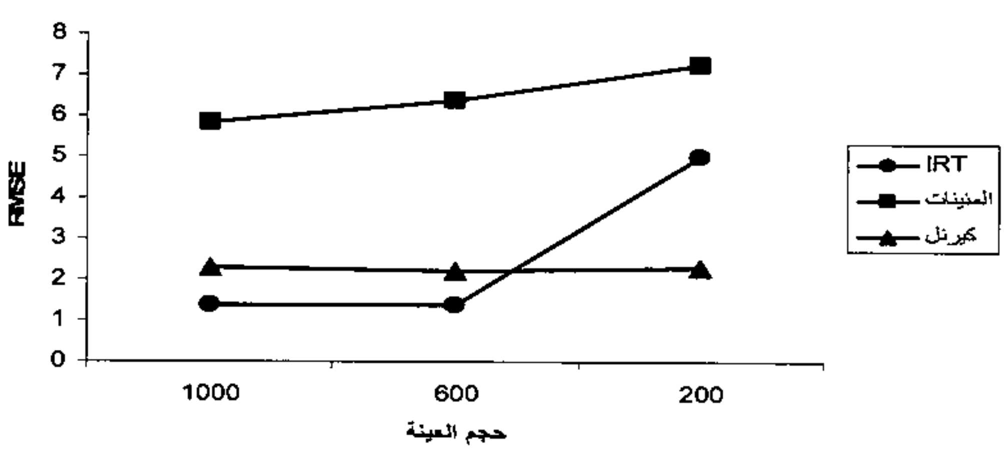
جدول رقم (١٦)

قيم (RMSE) لطرق المعادلة الثلاث للاختبار الذي تتالف من (٦٠) فقرة و المختلفة في الصعوبة

	200	600	1000
(IRT)	5.01932	1.3672	1.37868
المئينات	7.2481	6.3613	5.8265
كيرنيل	2.27309	2.19522	2.29115

يلاحظ من الجدول السابق ان اصغر قيمة من قيم (RMSE) كانت عند استخدام طريقة (IRT) ، في حين بلغت اعلى قيمة من قيم (RMSE) عند استخدام طريقة المئينات في معادلة الدرجات وعبر الاحجام المختلفة للعينات.ويظهر الجدول ان قيم (RMSE) تقل بازدياد حجم العينة وترتفع بنقصان حجم العينة.

يوضح الشكل رقم (١٢) قيم (RMSE) لطريقة (IRT) في معادلة الدرجات وعبر الاحجام المختلفة للعينات.



شكل رقم (١٢) قيم (RMSE) للمعادلة لطرق المعادلة الثلاث للاختبار الذي يتالف من (٦٠) فقرة والمختلف في الصعوبة

يوضح الشكل السابق ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيم RMSE فيها اقل قيمة، ويلاحظ من الشكل السابق ان قيم RMSE تنخفض مع ازدياد حجم العينة، كما ان

المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة، وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيم RMSE

ثانيا طريقة المئينات في معادلة الدرجات :

الخطا المعياري للمعادلة

يظهر من الجدول رقم (١٥) قيم الخطا المعياري لطريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر الاحجام المختلفة للعينات لنماذج الاختبارالتي تتالف من (٦٠) فقرة والمختلفة في الصعوبة.

لقد اظهرت نتائج الجدول ان طريقة المئينات في معادلة الدرجات كانت اقل الطرق دقة في معادلة الدرجات مقارنة مع طريقة (IRT) وكيرنيل, كما اظهرت النتائج ان الخطا المعياري الناتج عن المعادلة المئينية بتاثر بحجم العينة, فكلما زادت حجم العينة قلت قيمة الخطا المعياري للمعادلة والعكس صحيح.

ويظهر الشكل رقم (١١) قيم الخطا المعياري للمعادلة المئينية على سلم الدرجات وعبر الاحجام المختلفة للعينات.حيث يوضح هذا الشكل ان طريقة (IRT) وعبر الأحجام المختلفة للعينات، كانت قيمة الخطا المعياري للمعادلة فيها اقل قيمة، وان قيمة الخطا المعياري للمعادلة تنخفض مع ازدياد حجم العينة.

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يوضح الجدول رقم (١٦) قيم (RMSE) لنتائج المعادلة المئينية والتي تظهر ان طريقة المعادلة المئينية كانت اقل الطرق دقة في معادلة الدرجات وعبر الاحجام المختلفة للعينات.كما اظهرت النتائج ان قيم (RMSE) تزداد بانخفاض حجم العينة وترتفع بنقصان حجم العينة.ويظهر الشكل رقم (١٢) قيم (RMSE) للمعادلة المئينية على سلم الدرجات وعبر الاحجام المختلفة للعينات. حيث يظهر الشكل ان طريقة المعادلة المئينية كانت اقل الطرق دقة في معادلة الدرجات وعبر الاحجام المختلفة للعينات.كما ايظهر الشكل ان قيم (RMSE) تزداد بانخفاض حجم العينة وترتفع بنقصان حجم العينة.ويظهر الشكل رقم (١٢) قيم (RMSE) للمعادلة.

ثالثا: طريقة كيرنيل في معادلة الدرجات

يوضح الجدول رقم(١٥) قيم الخطأ المعياري للمعادلة لطريقة كرينيل في معادلة الدرجات على سلم الدرجات وعبر ألاحجام المختلفة للعينات. لقد أظهرت النتائج أن طريقة كرينيل في المعادلة كانت أفضل من طريقة المينات, كما ان حجم الخطأ المعياري للمعادلة في معادلة كيرنيل يعتمد على حجم العينة زيادة أو نقصانا. ويوضح الشكل رقم (١١) قيم الخطأ المعياري للمعادلة باستخدام طريقة كيرنيل على سلم الدرجات وعبر الأحجام المختلفة للعينات.

يوضح الشكل رقم(١١) ان طريقة كيرنيل في المعادلة كانت أفضل من طريقة المينات, كما يظهر الشكل أن حجم الخطأ المعياري للمعادلة في معادلة كيرنيل يعتمد على حجم العينة زيادة أو نقصانا.

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يظهر الجدول رقم (١٦) قيم (RMSE) لنتائج المعادلة باستخدام كيرنيل وعبر الأحجام المختلفة للعينات , حيث كانت طريقة (IRT) من أكثر الطرق دقة في معادلة الدرجات وعبر الأحجام المختلفة للعينات، في حين كانت طريقة كيرنيل أفضل من طريقة المئينات وعبر الأحجام المختلفة للعينات ،ويظهر الشكل رقم (١٢) قيم (RMSE) للمعادلة باستخدام طريقة كيرنيل على سلم الدرجات وعبر الأحجام المختلفة للعينات.

يوضح الشكل السابق ان طريقة (IRT) وعبر الأحجام المختلفة للعينات كانت قيم RMSE فيها اقل قيمة ويلاحظ من الشكل السابق ان قيم RMSE تنخفض مع ازدياد حجم العينة كما يلاحظ ان المنحنيات تميل الى النزول وتقترب من الصفرعندما ترتفع حجم العينة وهذا يشير الى ان ارتفاع حجم العينة يقلل من قيم RMSE.

مقارنة نتائج المعادلة للطرق الثلاثة للاختبار الذي يتالف من (٦٠) فقرة والمختلف في الصعوبة

الخطأ المعياري للمعادلة (SEE)

-تعتبر طريقة (IRT) من أكثر طرق المعادلة المستخدمة في هذه الدراسة دقة على سلم الدرجات وعبر الأحجام المختلفة للعينات

- تعتبر طريقة المئينات من أقل طرق المعادلة المستخدمة في هذه الدراسة دقة على سلم الدرجات وعبر الأحجام المختلفة للعينات

- تتأثر قيمة الخطأ المعياري للمعادلة بحجم العينة , فكلما زادت حجم العينة قل الخطأ المعيار للمعادلة, وعندما تقل حجم العينة يزداد الخطأ المعياري للمعادلة .

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

يظهر الجدول رقم (١٦) أن طريقة (IRT) من أكثر الطرق المعادلة المستخدمة في هذه الدراسة دقة على سلم الدرجات وعبر الأحجام المختلفة للعينات، كما يظهر الجدول رقم(١٦) أن قيم (RMSE) تزداد مع إنخفاض حجم العينة، وتقل بازدياد حجم العينة.

وفيما يلي ملخص لنتائج معادلة درجات الاختبار الذي يتكون من (٦٠) فقرة

جدول رقم (١٧)

ملخص نتائج طرق المعادلة الثلاثة لنماذج الاختبار الذي يتكون من (٤٠) فقرة

طريقة المعادلة	التشابه في الصعوبة						الاختلاف في الصعوبة					
	200		600		1000		200		600		1000	
	SE	RMSE	SE	RMSE	SE	RMSE	SE	RMSE	SE	RMSE	SE	RMSE
IRT	٢	٢	٢	٣	٢	٣	٢	٢	١	١	١	١
المئينات	٣	٣	٣	٢	٣	٢	٣	٣	٣	٣	٣	٢
كيرنيل	١	١	١	١	١	١	١	١	٢	٢	٢	٣

مناقشة النتائج والتوصيات

هدفت هذه الدراسة إلى مقارنة طرق كيرنيل والمئينات وطرق نظرية استجابة الفقرة عند استخدام تصميم الفقرات المشتركة في دقة معادلة درجات الاختبارات متعددة الحدود, وذلك من خلال الإجابة عن السؤالين الأتيين:

١- كيف يختلف الخطأ المعياري للمعادلة(SEE) عند النقاط المختلفة على سلم الدرجات تحت كل ظرف من :طول الإختيار, حجم العينة، والتشابه في مستويات الصعوبة؟

٢- كيف يختلف الجذر التربيعي لمتوسط الخطأ (RMSE) للمعادلة عند النقاط المختلفة عن سلم الدرجات تحت كل ظرف من: طول الاختيار، حجم العينة، وتشابه مستويات الصعوبة.

مقارنة نتائج طرق المعادلة الثلاثة من خلال حجم العينة:

الخطأ المعياري للمعادلة

أظهرت نتائج الدراسة تحت ظرف طول الاختبار والتشابه في مستوى صعوبة النماذج، ان نتائج طريقة المعادلة باستخدام (IRT)، وطريقة المعادلة باستخدام كيرنيل، قد أظهرت قيما منخفضة للخطأ المعياري مقارنة مع طريقة المعادلة باستخدام المئينات وعبر ألاحجام المختلفة للعينات.

كما أظهرت النتائج أن حجم العينة يؤثر في قيمة الخطأ المعياري للمعادلة, فكلما زادت حجم العينة قلت قيمة الخطأ المعياري للمعادلة, وإذا قلت حجم العينة زادت قيمة الخطأ المعياري للمعادلة.

ان الأختلاف في قيمة الخطأ المعياري للمعادلة باستخدام الطرق الثلاث المستخدمة في هذه الدراسة (IRT- المئينات – كيرنيل) يقل تدريجياً عندما يرتفع حجم العينة، كما أظهرت النتائج أن طريقة (IRT) كانت أكثر الطرق دقة مقارنة مع طريقة كيرنيل والمئينات وعبر اختلاف أحجام العينات.

إن مستوى صعوبة نماذج الاختبار لم يؤثر على حجم الخطأ المعياري للمعادلة بشكل كبير وخصوصا عند ارتفاع حجم العينة, بينما نجد أن طول الأختيار يؤثر على الخطأ المعياري للمعادلة، فعلى سبيل المثال عندما إرتفعت حجم العينة من ٦٠٠ –١٠٠٠ ،فإن قيمة الخطأ المعياري للمعادلة باستخدام طريقة كيرنيل عندما يكون الاختبار مكون من (٣٠) فقرة ومتشابه في مستوى الصعوبة إنخفضت من (٢،٧٠١٧) إلى (٢،٤٧٩٦).

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

تحت كل ظرف من طول الاختبار والتشابه في مستوى الصعوبة ، فان النتائج اظهرت تقارب نتائج (RMSE) مع نتائج (SEE) عندما ترتفع حجم العينة.

لقد اظهرت نتائج الطرق الثلاث ان نتائج قيم (RMSE) لكل طريقة يتاثر وبدرجة كبيرة بحجم العينة، فالعينات الكبيرة تنتج قيما صغيرة ل (RMSE)، والعينات الصغيرة تنتج قيما كبيرة لقيم (RMSE)، كما اظهرت النتائج ان قيم (RMSE) تقل في الطرق المختلفة مع ازدياد حجم العينات.

كذلك اظهرت قيم (RMSE) الناتجة عن تطبيق الطرق الثلاث في معادلة الدرجات ،ان طريقة (IRT) في معادلة الدرجات كانت اكثر الطرق دقة في معادلة الدرجات على سلم الدرجات وعبر الاحجام المختلفة للعينات. كما دلت النتائج ان طريقة كيرنيل في المعادلة كانت افضل من طريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر الاحجام المختلفة للعينات.

كما اظهرت نتائج هذه الدراسة ان الإختلاف في صعوبة نماذج الاختبار لا يؤثر بدرجة كبيرة في العلاقة بين حجم العينة وقيم (RMSE) وهذا ما اظهرته نتائج هذه الدراسة.

مقارنة نتائج طرق المعادلة الثلاث من خلال طول الاختبار

الخطا المعياري للمعادلة

تحت الظروف المختلفة لحجم العينة والتشابه في صعوبة النماذج، اظهرت النتائج ان طرق المعادلة باستخدام طريقة (IRT) وكيرنيل اظهرت قيما منخفضة للخطا المعياري مقارنة مع طريقة المئينات وعبر الاحجام المختلفة للعينات.

ان طول الاختبار يوثر في الخطا المعياري للمعادلة، فالاختبار الطويل يعطي قيما كبيرة للخطا المعياري للمعادلة، على حين ان تقليل طول الاختبار من (٦٠) الى (٤٠) يؤدي الى تخفيض كبير للخطا المعياري للمعادلة مقارنة مع التخفيض عندما يكون طول الاختبار يتناقص من (٤٠) الى (٢٠).

فعلى سبيل المثال في طريقة (IRT) عندما كان الخطا المعياري (3,42516) للاختبار الذي يتكون من (٦٠) فقرة والمتشابة في صعوبة النماذج وعند حجم عينة (١٠٠٠) ، فان هذا الخطا

المعياري ينخفض الى (1,49027) للاختبار الذي يتكون من (٤٠) فقرة وبنفس حجم العينة والمتشابه في صعوبة النماذج.

ان الاختلاف في قيم الخطا المعياري للمعادلة باستخدام الطرق الثلاثة يصبح اقل عندما ينخفض طول الاختبار، ان طريقة (IRT) كانت اكثر الطرق دقة في هذه الدراسة مقارنة مع الطرق الاولى عند استخدام معيار الخطا المعياري للمعادلة.

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

ان نتائج قيم (RMSE) تتشابه مع نتائج (SEE) للطرق الثلاث المستخدمة في هذه الدراسة وعبر الظروف المختلفة من حجم العينة والتشابه في صعوبة النماذج، ويمكن تلخيص اهم النتائج كما ياتي :

١-دلت نتائج قيم (RMSE) الناتجة عن استخدام الطرق الثلاث في المعادلة ان طريقة (IRT) وكيرنيل اظهرت قيما صغيرة ل (RMSE) مقارنة مع طريقة المئينات.

٢-ان زيادة طول الاختبار يؤدي الى ارتفاع قيم (RMSE).

٣-ان الاختلاف في قيم (RMSE) و عبر الطرق الثلاث يقل مع زيادة طول الاختبار.

٤-تعتبر طريقة (IRT) من اكثر الطرق دقة في معادلة الدرجات مقارنة مع طريقة كيرنيل و طريقة المئينات في معادلة الدرجات على سلم الدرجات وعبر الاحجام المختلفة للعينات.

مقارنة نتائج طرق معادلة الثلاث من خلال التشابه في صعوبة الفقرات

الخطأ المعياري للمعادلة

تحت الظروف المختلفة لحجم العينة والتشابه في صعوبة النماذج، اظهرت النتائج ان طريقة المعادلة باستخدام (IRT) وكيرنيل اظهرت قيما منخفضة للخطا المعياري مقارنة مع طريقة المئينات وعبر الاحجام المختلفة للعينات.

ان الاختلاف في قيم الخطا المعياري للمعادلة للطرق المعادلة الثلاثة يميل الى الانخفاض قليلا عندما تكون النماذج مختلفة في مستوى صعوبتها.

ان قيم الخطا المعياري لطرق المعادلة الثلاثة لنماذج الاختبار الذي يتكون من (٤٠) و (٦٠) فقرة تنخفض عندما تختلف النماذج في مستوى صعوبتها، فالنماذج المختلفة في صعوبتها تميل قيم الخطا المعياري الى الانخفاض عندما تختلف مستويات الصعوبة فيها.

الجذر التربيعي لمتوسط مربعات الخطا(RMSE)

تحت الظروف المختلفة من طول الاختبار وحجم العينة، اظهرت النتائج ان طرق المعادلة باستخدام (IRT) وكيرنيل اظهرت قيما منخفضة ل (RMSE) مقارنة مع طريقة المئينات وذلك بسبب الاختلاف في صعوبة النماذج للاختبار .

لنماذج الاختبار إلذي يتكون من (٦٠) فقرة فان قيم (RMSE) الناتجة عن تطبيق طرق المعادلة الثلاث انخفضت وللطرق الثلاث عندما اختلفت صعوبة النماذج، كذلك للاختبار الذي يتكون من (٤٠) فقرة فان قيم (RMSE) للطرق الثلاث تميل الى الانخفاض اما نموذج الاختبار الذي يتكون من (٣٠) فقرة فانه لا يوجد له نمط محدد يمكن من خلاله تحديد قيم (RMSE) فيما اذا كانت ارتفعت ام انخفضت نتيجة لاختلاف صعوبة نماذج الاختبار.

وفي مجال مقارنة نتائج هذه الدراسة مع الدراسات الأخرى, فقد إتفقت نتائج هذه الدراسة مع دراسة(Wang, 2005) التي هدفت إلى مقارنة الطريقة اللوغارتمية الخطية المستمرة مع طريقتين في المعادلة: طريقة كيرنيل وطريقة المعادلة المئينية حيث اظهرت نتائج الدراسة أن طريقة كيرنيل كانت أكثر دقة من طريقة المعادلة الميئنية.

كما اتفقت نتائج هذه الدراسة مع دراسة ليفنجستون (livingston,1993) التي هدفت إلى فحص دقة المعادلة باستخدام طريقة كيرنيل مع أحجام عينات صغيرة وباستخدام تصميم المجموعات غير المتكافئة ذات الاختبار المشترك ،حيث أشارت النتائج الى أن طريقة المعادلة باستخدام طريقة كيرنيل كانت أكثر دقة من المعادلة باستخدام طريقة المئينات.

كما اتفقت نتائج هذه الدراسة مع دراسة (Yuki,2008) والتي أشارت إلى أن طريقة IRTكانت أكثر دقة في معادلة الدرجات من طريقة كيرنيل . كما اتفقت نتائج هذه الدراسة مع دراسة يونغ وو (YOUNG WOO,2007) والتي أشارت إلى ان طريقة IRT كانت أكثر دقة في معادلة الدرجات من طرق المئنات . في حين نجد أن نتائج هذه الدراسة قد اختلفت مع نتائج دراسة أيوب (١٩٩٤) والتي أشارت إلى أن طريقة المئينات كانت أكثر دقة في معادلة الدرجات من طريقة IRT ،كما اختلفت هذه الدراسة مع نتائج دراسة أماندا (Amanda,2008) التي أظهرت أن طريقة الميئنات في المعادلة كانت أكثر دقة من طريقة IRT.

الاستنتاجات

هدفت هذه الدراسة إلى مقارنة طرق كيرنيل والمئينات وطرق نظرية استجابة الفقرة عنداستخدام تصميم الفقرات المشتركة في دقة معادلة درجات الاختبارات متعددة الحدود , وذلك من خلال الإجابة عن السؤالين الأتيين:

١- كيف يختلف الخطأ المعياري للمعادلة(SEE) عند النقاط المختلفة على سلم الدرجات تحت كل ظرف من :طول الإختيار , حجم العينة والتشابه مستويات الصعوبة؟

٢- كيف يختلف الجذر التربيعي لمتوسط الخطأ (RMSE) للمعادلة عند النقاط المختلفة عن سلم الدرجات تحت كل ظرف من : طول الإختيار ، حجم العينة، وتشابه مستويات الصعوبة.

وقد خلصت الدراسة الى الاستنتاجات الآتية:

١-بصورة عامة , توصلت هذه الدراسة إلى أن طريقة IRT وطريقة كيرنيل في معادلة الدرجات كانت أكثر دقة من طريقة المئينات في المعادلة وعبر جميع المتغيرات التابعة التي استخدمت في هذه الدراسة .

٢-أظهرت نتائج هذه الدراسة أن طريقة IRT في المعادلة كانت أكثر الطرق دقة مقارنة بطريقة المئينات وكرينيل وعبر جميع المتغيرات التابعة المستخدمة في هذه الدراسة.

٣-ان طريقة IRT وطريقة كيرينل تميل قيم الخطأ المعياري فيها للإنخفاض مع إرتفاع حجم العينة وخصوصا لنماذج الإختبار من (٤٠) -(٦٠) فقرة، كذلك فإن كلا الطريقتين تميل قيمRMSE فيها إلى الانخفاض مع ازدياد حجم العينة.

٤-بصورة عامة أظهرت نتائج هذه الدراسة أن طريقة IRT وكرينيل أظهرت وبصورة تقريبية نفس النتائج عندما إختلفت مستويات صعوبة النماذج.

٥- أظهرت نتائج هذه الدراسة أن طريقة كيرنيل كانت أكثر دقة من طريقة المئينات ، حيث دلت النتائج على أن طريقة المئينات كانت أقل الطرق دقة مقارنة مع طريقة IRT و كيرنيل.

٦- أظهرت نتائج هذه الدراسة أن أحجام العينات الكبيرة يوثر على معايير دقة المعادلة, فحجم العينات الكبير يؤدي إلى تقليل الخطأ المعياري,SEE وتقليل قيم RMSE.

٧- بصورة عامة فإن الإختلاف في الخطأ المعياري للمعادلة وقيم RMSE يصبح أقل عندما ترتفع حجم العينة.

٨- الإختيار الطويل يؤدي إلى ازدياد قيم RMSE وهذه يمكن تفسيرها في عاملين رئيسين:

أ-ان قيم RMSE المستخدمة هنا ليست قيما معياري (Standardized) لأن فقرات الاختبارات لم تقسم باستخدام تطابق الإنحراف المعياري للعلامات المشاهدة ولأن الإنحراف المعياري للدرجات الصحيحة يرتفع مع زيادة طول الإختيار ، فانه يتوقع أن القيم المعيارية لقيم RMSEسوف تظهر ميلا قليلا إلى الارتفاع مع ازدياد طول الإختبار.

ب- ان الإختبار الطويل يعني إحتمالية عالية عند نقاط الدرجة وبالتالي مفحوصين قليلين لكل درجة للعينات المركبة.

ان العدد القليل من المفحوصين عند كل درجة يقلل من دقة المعادلة عند كل النقاط.

التوصيات

١- عدم استخدام الطريقة المئينية في معادلة الدرجات متعددة الحدود.

٢- استخدم الباحث لتقييم إجراءات المعادلة معياري الخطأ المعياري وجذر متوسط مربعات الفروق، لذلك يوصي الباحث باستخدام معايير أخرى مثل الصدق التقاطعي ومعيار الاهمية النسبية للمعادلة.

٣- يوصي الباحث بتطبيق الطرق الثلاث على بيانات حقيقية متعددة الاستجابة , وذلك للوقوف على مقدار الإختلاف في المعادلة بين البيانات الحقيقية والبيانات التجريبية.

٤- استخدام تصميم آخر غير تصميم الفقرات المشتركة للوقوف على مقدار الإختلاف في دقة المعادلة لنفس الطرق المستخدمة في هذه الدراسة.

المراجع

اولاً: المراجع العربية:

الحربي،عيسى جود الله حميد (٢٠٠٩). **اثر تمثيل الفقرات للمحتوى ونسبة الفقرات المشتركة وطرق توزين الفقرات على دقة معادلة درجات الصور للاختبار عندما يكون التصميم المستخدم هو الجذع المشترك (الفقرات المشتركة).** اطروحة دكتوراة غير منشورة ، الجامعة الاردنية ، الاردن.

الدوسري، راشد حماد (٢٠٠٤). **القياس والتقويم التربوي الحديث: مباديء وتطبيقات وقضايا معاصرة.** دار الفكر، عمان.

الصمادي، إسماعيل محمد محمود،(٢٠٠٧). **أثر طريقة اختيار الفقرات في اختبار الجذع المشترك على دقة معادلة اختبار متعدد المستوى في الرياضيات للمرحلة الأساسية في الأردن،** رسالة دكتوراه غير منشوره ،جامعة عمان العربية، عمان، الأردن.

الصمادي، مروان صالح علي، (٢٠٠٦). **فاعلية طرق تصحيح اختبار الصواب – الخطأ المتعدد وتأثيرها على دقة معادلة الاختبار باستخدام نماذج النظرية الحديثة للقياس،** رسالة دكتوراه غير منشوره ،جامعة اليرموك ،عمان ، الأردن.

المدانات، رائد فايز، (٢٠٠٨). **أثر طريقة المعادلة باستخدام جذع مشترك وعدد فقراته وحجم العينة على القيم المعادلة والخطأ في المعادلة بين صورتي اختبار في الفيزياء،** رسالة دكتوراه غير منشوره ،جامعة عمان العربية،عمان ، الأردن.

أيوب، حسين محمد عبد القادر، (١٩٩٤). **المقارنة بين أربع طرق للمعادلة عندما يكون تصميم من مجموعات متكافئة وغير متكافئة،** رسالة دكتوراه غير منشوره ،الجامعة الاردنيه ،عمان، الأردن.

ثانياً: المراجع الأجنبية

Akour, Mutasem Mohammad Mahmoud. (2006). **A comparison Of various equipercential and Kernel equating methods under the random groups**. Unpublished Doctoral Dissertation, University of Iowa,USA.

Alina A, Von Davier, Paul W. Holland, Dorothy T.Tayer. (2004).**The Kernel method of test equating.** New York: Springer – Verlag.

Amanda A, Walkouts. (2008). **A comparison of classical test theory and item response theory methods for equating Number-right scored to formula scored assessments** .Unpublished Doctoral Dissertation, University of Kansas, USA.

Angoff, W.H. (1971). **Scales, norms, and equivalent scores**.InR L.Thorndike (Ed), Educational measurement (2nd ed.) (pp.508-600).Washington: American Council on Education.

Baker, F. B., & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. ***Journal of Educationa Measurement, 28*, 147-162.**

Cook,L.L&Eignor,D.R.(1991).An NCME instructional module on IRTequating methods. Educational Measurement :**Issue and Practice,10,37-45.**

Embreston.S.&Reiase,S.(2000).**Item Response Response Theory for Psychologists**.NewJersey: Lawrence Erlbaum Associates,Inc.

Jinghua, Low Albert C. (2008). A comparison Of the Kernel Equating Method with Traditional Equating Methods Using SAT®Data. **Journal of Educational Measurement**.45 (4),309-320.

Hambelton,R & swaminthan,H. (1985). **Item Response Theory: Principles and Applications.** Boston, Kluwer: Nijhoff Publishing.

Hambleton, Linden. (1982). Advances in Item Response Theory and Applications: An Introduction. ***Applied Psychological*** **Measurement**; 6, 373-378.

Hambelton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals* of item response theory. New York: Sage Publications.

Han, K.T. Hambleton, R.K, **User manual for winGen: Windows software that generates (IRT) model parameters and item response**. New York: Holt Rinehart and Weston.

Harris D.J. &, Kolen, M. J. (1990). Effect of examinee group on equating relationships. **Applied Psychological Measurement,** 10,35-43.

Harwell, M.et, al. (1997).Analyzing the Results of Monte Carlo Studies in Item Response Theory. **Applied Psychological Measurement,** 57(2), 266-280.

Holland, P.W.Thayer, D.T. (1987). Notes on the use of log-linear models for fitting discrete probability distributions. **(Technical Report 87- 79), Princeton, NJ: Educational Testing Service**

Hue, Geo. (2004).**The effect of different anchor tests on the accuracy of test equating for test adaptation.** Unpublished Doctoral Dissertation, Ohio University, USA.

Huigin, Hu. (2004). **Investigation of (IRT)-based equating methods in the presence of outliers**. Unpublished Doctoral Dissertation. University of Alberta, Canada.

Kolen, M., J. (1981). Comparison of traditional and item Response theory methods for equating tests. **Journal of Educational Measurement,** 18 (1), 1-11.

Kolen, M., J. (1988).Traditional equating methodology. **Educational Measurement: Issues and Practice,**7,29-36.

Kolen, M. 1 & Brennan, R. L. (2004). **Test equating, scaling, and linking and practices**, (2^{nd} ed). New York: Springer- Verlag.

Lee, G.Hill, Kolen, M. J., Frisbee, D. & Ankenmann, R. (2001). Comparison of dichotomous and polytomous item Response models in equating scores from tests composed of test lets. **Applied Psychological Measurement,** 25 (4), 357-372

Lord, D. M. (1980). **Applications of item Response Theory to Practical Testing Problems** Hillsdale. N.J: Erlbaum.

Mao, X. (2006). **An investigation of the accuracy of the estimates of standard errors for the kernel equating functions.** Unpublished Doctoral Dissertation, University of Iowa,USA.

Peterson, N. S, kolen , M J , & Hoover , H. D. (1989).Scaling ,Norming and Equating. **Educational measurement, Washington D.C: American Council on Education**: 241-262.

Qu, Yanxuan. (2007). **The Effect Of Weighting In Kernel Equating Using Counter-Balanced Designs**. Unpublished Doctoral Dissertation, Michigan State University, USA.

Robert R, Keller. (2007). **A comparison of item response Theory true score equating and item response theory- Based local equating.** Unpublished Doctoral Dissertation, University of Massachusetts, USA.

Skaggs, G. & Lissitz, R. W. (1986). An exploration of the robustness of four test equating models. ***Applied Psychological Measurement, 10*, 303-317.**

Young Woo, Cho. (2007). Comparison of bootstrap standard errors of equating using (IRT) and equipercentile methods with polytomously-scored items under the common-item Nonequivalent- groups design. Unpublished Doctoral Dissertation, University of Iowa.USA

Yuki, Nozawa. (2008). **Comparison of parametric and Nonparametric (IRT) equating methods under the common- item nonequivalent groups design**. Unpublished Doctoral Dissertation, University of Iowa.USA

Victor K. (2007). Equating accuracy using small samples in the random groups design . Unpublished Doctoral Dissertation, Ohio University, USA.

Wang, T. (2005a). An alternative continuization method to the Kernel method in von Davier, Holland and Thayer's (2004) test equating framework. **(CASMA Research Report 11).** Iowa City: Center for Advanced Studies in Measurement and assessment.

A Comparison of Kernel , Equipercentile Methods, and IRT Methods Under Anchor Items Design With Respect to Accuracy Of Equating Polynomial Scores

by

Yousef Abed Alaty Almahrouk

Supervisor

Dr. Khalil M. Olian

ABSTRACT

The objective of this study is to make a comparison between Kernel, Equipercentile Methods, with IRT Methods Under Anchor Items esign with Respect to Accuracy Of Equating Polynomial Scores by tudying the following variables:

1 - Sample size: three levels used for the sample size :: 200, 600.1000, and the sizes of the samples are suitable for the equation using the methods of random groups design.

2 - Test length: three levels used for test length :30-40-60 items

3 - Difficulty levels: Used two levels of difficulty: difficult similarities and the differences in difficulty (difficulty difference = 0.01 standard deviation) .

It has been reached to the results of the accuracy of the equating in the light of the previous variables that are used individually and collectively, where the accuracy of the equating were estimated

using two standard: the standard error of the equating, and the square root of the mean error.

This study specifically tried to answer the two questions below :

1- How is the standard error of the equating different at different points on the scale of grades under all circumstances: test length, sample size, and the similarity of levels of difficulty?

2- How is the standard residuals of the equating different at different points on the scale of grades under all circumstances: test length, sample size, and the similarity of levels of difficulty?

this study Choose Kernel, Equipercentile Methods, and IRT Methods because it is more common methods in equating; to detect the accuracy of equating tests using random groups design with common items.

In order to compare the three methods of equating, we produced trial data using software tool - wingen 2: by using three variables:test length, sample size, and the similarity in the levels of difficulty. by using item response theory to generate responses of items of the test, where the equivalent test scores with using scores Observation in the theory of item response as the main criterion of the equating, as well as the use of the standard error of the equation residuals and the standard for assessing the results of the equation.

The results indicated that the method of the equating using the item response theory was the most accurate, which was followed by Kernel method, and in the various conditions that have been addressed in this study, the results also showed that the large sample size reduces the standard error of the equating and reduces the residuals of the standard.